

# APPLICATION OF BUSINESS ANALYTICAL PROCEDURES

## Harshit Dasgupta

A31506422098

Bachelor of Business Administration – International Business (Class of 2022-2025)  
Amity Global Business School, Pune – 411016, Maharashtra, India.

### I. INTRODUCTION

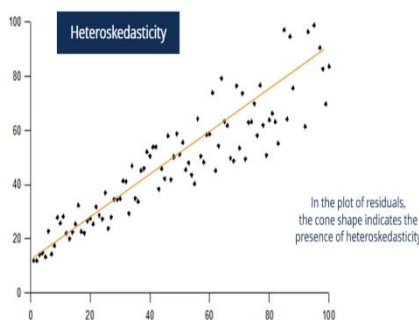
Datum or Data is a key incentive as well as a concern, for any business. Data will incentivize a business, to invest more in collecting raw data and analyze it further using various techniques. Export Transformation and Loading is a key activity, to derive sense out of the data collection. Consider an example, ABC Company wanted to inject another product into the product assortment width, Fundtech. Datum would have incentivized ABC Company to, collect reliable (preferably from a primary source) in the pilot study. Google Analytics is an effective tool, for measuring KPIs and tracking the entire life cycle, ranging from Stickiness Rate to Organic search measurements. This will enable ABC Company to either do an A/B test or to use a new web service as a prototype, to reach wider audiences lengthwise.

The data collected and organized in an analytical-friendly model can then be subjected to various statistical and econometric techniques. For instance, heteroskedasticity can be quantified by the nature of values and constraints values.

Heteroskedasticity of a simple linear regression<sup>1</sup>

Heteroskedasticity is a serious problem, as a regression using Ordinary Least Squares assumes a constant variance of the residuals drawn from the population of sample data. Further, Pure or Impure heteroskedasticity can be categorized, which will help eliminate anomalies, which we will discuss in future pages.

Figure 1



#### Heteroskedasticity Applications

Datum will also involve predictive analysis. Predictive analysis enables a firm to discern and analyze the patterns of past and present records, then simulate the results while assuming constraints to stay similar to present market conditions in the future. This uses Binary inputs, such as Yes or No. A common example of an application is Decision Tree

analysis. Predictive analysis however can also be done by categorizing data based on similar categories. For instance, Suppliers can be grouped into groups, based on punctuality using Expectation-Maximization. Furthermore, Time series prediction analysis techniques also exist which commonly can be achieved using ARIMA, Moving Average and Infitum.

Consider an example of Pick Point as the exact cause of lower profits in a particular region. After analyzing sales data, based on regions, viz-a-viz West, south, or east, we were able to discern the probability of profits from every region. This probability would depict, a key aspect that causes lower sales in a particular region. Thus, probability acts as an intrinsic strategy more than a holistic strategy. This helps us to market research, on the identified areas, for instance, potential demotivators for customers from a brand service, with limited expansions, compared with colossal FinTech Giants. Thus, market research insights will assist in undertaking key decisions and identifying the scope of improvements. The key insights will be discussed in detail in this report further.

Data as a concern, basically refers to excessive flow of information. Data that is not hashed may be available to your competitors and thus disrupt the chance of a firm having its Unique Value Proposition or USPs. Data also makes a business vulnerable to false claims thus converting their bullish market to a bearish tone.

Another aspect of concern is the unification of data. Multiple data genres such as String, Numeric, sequence, Boolean Genre, and Infinite have to be managed separately. These data cannot be grouped or may not always work with each other. For example, using the list as your x-axis and numerical data as your y-axis will disrupt your results, compared to what data or string and numerical information may possess.

#### 1.1 About The Company

FinTech is a portmanteau for Financial Technology. Currently, FinTech companies account for >14% CAGR with the market size being \$312.94 Billion. ABC Company is a bootstrapped company, that was founded in 2020 with a vision of unifying multiple asset-class investments into a single platform; It supports an automated system, making it easier and affordable to provide investment knowledge, expertise, assistance, and advisory through multiple marketing channels, for larger sections of society.

The company's target segments are industries in the automobile sector, auditing, and advisory services. Typically, it focuses on B2B services more than B2C services. The ABC is recognized by BSE, NSE, And start-up India. Along with funds, the room is also rewarded with RERA AND AFMI Certification from the Skill India initiative.

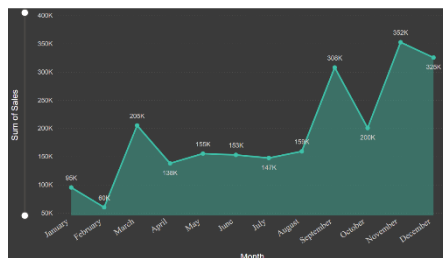
<sup>1</sup> Source: [Heteroskedasticity - Overview, Causes, and Real-World Example \(corporatefinanceinstitute.com\)](https://www.corporatefinanceinstitute.com)

**Table 1**  
Percentage Sales OF ABC Company Product Mix (2020)

Product Mix	Percentage Sales in Qtr.	
	1/2	3/4
Funds Pay	32.31 %	33.84 %
Funds Tech	32.10 %	29.74 %
Funds Research	35.59 %	36.42 %

Currently, the CEO and co-founder of ABC Company is Masroor Khan. The firm served 200,000 customers in 2024, consisting of 40,000 investors as customers. The customer base is estimated to rise to 500,000 in 2025. ABC Company is headquartered in Pune, East Street Galleria 411001. The employee base consists of 50 team leaders and 51-200 employees. 138 members and more than 800 interns are associated with ABC Company.

Figure 2



ABC Company sales for the year 2020

The company had an Assets Under Management (AUM) of \$12 Million in 2024, with its current valuation being \$1.8 Million (2024). Asset under management (AUM) is projected to reach \$60 Million, a compound annual growth rate of 5% and a Return on Investment of 23% in 2024. The company has a 0% equity share and thus enjoys a full 100 % stake (Sweat Equity Vesting) in the market, thus not maintaining ramen profitability.

The product mix, of ABC Company currently in the market are:

1. ABCPay: Investment products with personalized strategic decisions through a wealth management system, with the motive to safeguard a family’s future. The target segment is a typical middle-class family.
2. ABCTech: A platform for pocket-friendly customized web and app development interfaces for small businesses that contain Technological products like auditing, autopayment, MF, etc.
3. ABCAudit: Investment advisory Board and knowledge center that simplifies the tiresome process of audit inspections, with features like auto-audit and data-driven cost analytics.
4. ABCResearch: This is the research wing that consists of experienced financial advisors who guide customers to accurate financial investment decisions in multi-asset allocations.

ABC Company enjoys strategic partnerships with FinTech Brands such as Paytm and AngelOne and other firms such as Motilal Oswal Investment Services. Overall, the company’s target segment involves high-net-worth individuals and retail investors. In a nutshell, ABC Company is ‘Investment Ki Dukaan’!

The competitive landscape signifies the firm’s threats that exist outside the business boundaries. By glancing at the below figure, we can discern that ABC Company belongs to the performer’s category, usually involving firms with an inspirational ability to skyrocket their market share over time. This means that though not perfect, ABC Company has the potential to grow its market share. Market share doesn’t only grow using traditional models such as advertisements but also modern techniques.

In-game advertising is a trend, used to target younger segments of the audience depending on the genre of the game. Deloitte for instance belongs to Leaders, behind

IBM Consulting due to their satisfaction level. This makes the customer subscribe to the Financial Services, even when a customer may have a bad customer service experience twice or thrice.

1.2 Objectives of the Study

The objectives for the study, undertaken are as follows:

- A. To Discern the extraneous areas concerning flag-bearing cautions in gross sales using data shreds of evidence using magnitude ( $\Delta$ )
- B. To find statistical advancements to detect risks and risk perception by stakeholders.
- C. To analyze countless attributes acting as a hurdle or determinant in an embryonic firm compared to an established one.

1.3 Context Of the Study

Financial risk management is undergoing a seismic shift, driven by the transformative power of colossal data analytics; Financial institutions are now leveraging vast datasets not just as historical records but as powerful tools to revolutionize risk management practices [Olaiya, Obani et al]. This explains the study’s inference on business processes and elements using a solid database. A FinTech service-providing firm is at constant risk of outsider elements. Unpredictability and variability a very common risks a firm has to be bound to.

The theme of this study involves research of data provided by different sources, for different typecasts such as Sales, or even Predictions. These data sets will also be used, to efficiently develop the first step before the market survey could take place. After the market survey by the on-field marketing team, the data developed by them will be analyzed and developed into a dashboard acting as a Decision Support System. Thereafter the Executive support system will be developed, for the upper management to take the right decisions and precautions by using insights.<sup>2</sup>

The context also involves insider elements, as in the form of actual and predicted data. These datasets can be used to discern gross sales. The vast influx of information enables a more holistic and nuanced discerning of risk factors, moving beyond traditional metrics to incorporate a wide array of external variables [Usoor, Nwafor et al]. These metrics can be used to communicate insights, through Data Storytelling. For instance, in Power BI, Narrative Functions under AI Visuals Ribbon, Insert Tab is a tool that guides the executive to view and discern the dashboard visuals.

Ergo, mainly the setting is based on determining the market position, and picking the affected sales areas for the ABC Company. The illumination of the database depicts the seasonal variation in sales. As discussed *a priori*, quality of service such as delayed customer service and constant lag occurring in Transaction systems ad infinitum, affect the plunge in sales to the Treble.

II. LITERATURE REVIEW

Countless existing theories on Business Analytics and its application, are currently present. To discern Data measurement and the ultimate goal of making the right calls at the right time, we must discern the underlying patterns of data, ranging from data storage, handling NULL values, ETL Process ad infinitum.

<sup>2</sup> NB. The datasets (except the Market Research) contain skewed values, to maintain confidentiality. The calculations done have a high similarity index to the real dataset.

Data is only as good, as its effective utilization. For instance, if data is not handled as per company strategies, to do this, data must be identified.

$$D(q) = A + Bp(x)$$

As a layman, one will consider it to be a final equation and use it to forecast sales. However, we are not certain, whether this equation will

1.1 Database Concepts

Before we encounter a journey of decision-making, we must possess an accurate data source. The most popular data warehouse, where data is stored, is an Excel Sheet. However, when it comes to a more secure dataset (Two or more relations), nothing is better than the MySQL portmanteau of Structured Query Language Workbench, which is a graphical user interface for a database entity. A database management system or DBMS is considered an entity, used for designing, maintaining, and implementing database systems. A database is an effective tool especially when handling Colossal Data, as well as effective retrieval of Data.

A Database management system usually consists of multiple relations or tables. A collection of relations is a database. A relation matrix is nothing but an arrangement of fields and tuples, where every field has a well-defined value genre. For instance, a field may consist of a varchar, float, or integer. One can even define, the limitation of the value. A database can be defined, using a set of codes called languages. There exist multiple genres of languages in an RDBMS system. The genres of languages are mentioned below in detail:

A. Data Definition Language:

DDL refers to a genre of language, that involves defining the structure of a database schema. This language involves queries that are useful for the development of a framework of a database. For instance, Data definition language is involved in defining the foreign key, primary key, or default values. Data definition language can also define, Null or Not Null values. The statements, involved are CREATE, DROP, TRUNCATE, and Infinitum.

B. Data Manipulation Language:

DML refers to a genre of language *A posteriori*, which involves Queries that are used to exploit data from a relation, in a database. For instance, a Data manipulation language will be involved to INSERT, SELECT, UPDATE, DELETE etcetera in a logical database. The same will be considered statement queries. Data securitization is a key concern faced by Countless industries today, especially those dealing with Financial Liquidation sensitive information. LOCK query can be used to prevent access of others to a relation matrix. It relates to Database Object Creation.

C. Data Control Language

A DCL is a genre of language, that deals with parent and child relations. The statements used are GRANT and REVOKE. As the name suggests, grant command is used to grant prestigious rights to a particular user, typically coined as a parent. A child is a user that is dependent on the parent. Revoke command is the opposite. It is used to revoke the prestigious rights of a user.

D. Transaction Control Language

ATCL is used to maintain the overall objectivity and power of transactions. Typically, transactions unify a set of commands acting like a Holistic Rubric, defining the positions of the commands. Consider an example of the calculation of a new measure using a syntax. We can declare a new

variable temporarily, genre a select statement to retrieve the desired data, the end with COMMIT. Commit adds permanency to the syntax, which can only be eliminated using ROLLBACK. Therefore, it avoids accidental anomalies.

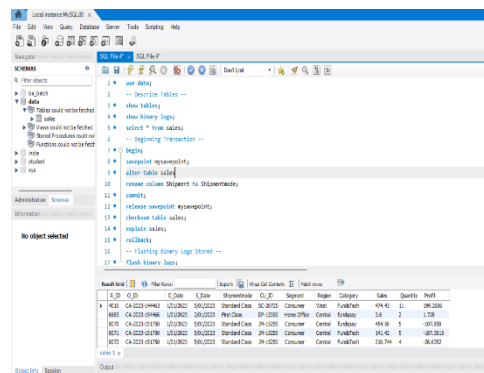
In a logical database, a server must define key constraints after the creation of data. For instance, MySQL doesn't have an undo function, which a typical office suite may pertain. Regardless, SAVEPOINT can be used, as a checkpoint system. This class statement can work with the local host SQL server as well as with dependency injectable files. MySQL cannot import a pure xl xs workbook, however, it can import comma-separated value (CSV) files. A CSV file is useful, as it is a plain text file, separating values with commas, inclusive of a tabular representation of data, typically found in an Excel sheet. To discern this, consider a relationship between 5 fields and 5000 tuples about sales. You can discern the relation schema, using the EXPLAIN Clause or DESCRIBE. After importing data from a CSV file using the table data import wizard, your genre a series of queries, to modify and retrieve data from a relation.

However, you want to use the commit clause to save the progress made above the statement for a particular Chunk of the dataset.

In the example mentioned above, sales data for annual 2023 warehoused in a relation can be explained using the desc sales, algorithm. This will enable the server to comprehend the Data GENRE entered while creating a relation, defining NOT NULL, and setting default constraints. Thereby, including suitable data genre algorithms, for a particular field. Furthermore, a super key for tuples or a primary key for fields can also be used, as per a firm's procedure. What field defines the key aspect of decision-making? What field helps to identify the relation as a holistic approach, are some of the questions, a user will have to answer while defining permissions.

To do so, you can use BEGIN and END statements. Begin and end act as an imitating and stop point respectively in a relation. This can help a user define transactions, drop variable transactions, declare variables et.al. Multi-layer securitization protocol is available on MySQL. To Solve this concern, Flush binary logs are created. To administer it on stored tables, a user can use SHOW statement, to list out the available relations. Data convertibility is a term, that refers to the genre of permission, that a dataset consists of. For instance, data permission may be READ only or may allow the user to modify or WRITE on the relation. To define these user rights, LOCK and UNLOCK statements are used. 'Lock sales read' here will disallow modification of data. 'Flush tables with read lock' will remove any permissions previously defined as READ Only.

Figure 3 Practical implementation in MySQL workbench 8.0



Every second spent on data retrieval is related to the key precious resource lost for the firm. This involves entering queries to retrieve the data, as well as the second spent for data retrieval. For example:

```
'SELECT Shipment mode, Segment where sales>200 GROUP BY shipment mode, ORDER BY sales;'
```

This statement will lead to an immense loss of time and effort. This parameter is the key, to defining a unified input (IN) or output (OUT). Before defining a parameter, one must use a delimiter \$\$ to chunk the queries written. It helps efficient use of the END statement when defining a parameter. Indexing is also useful for faster retrieval of data, though the data limit is set by the Graphical user interface. The practical algorithmic exchanges of these are given in the figure below.

Figure 4

Practical operation of indexing and procedure

Therefore, data stored in a database can be evaluated per users' requirements. The potential of the database is not limited to what was stated above. The database acts as a filter and a ground zero before the datum can be processed supplementally. Data correction and insights by logical operators along with handling data mesh can be performed using MySQL.

A chief concern now arises, how can the dataset be treated further, so that a firm can make operative decisions? The next section undertakes the concepts of quantifying qualitative aspects of data, while the measurement of sales data is voided.

1.2 Quantifying Qualitative Data

When confronted with the poor predictive performance of a model, it is usually a good idea to first inspect its input data [Lange, Biessmann, et al.]. The input data be it, quantitative or qualitative must undergo certain hyperparametric aspects of measurement. For instance, suppose, the Datum collected contains some NULL values portmanteau of Missing values. In this scenario, the results generated will be wrong insight, thus leading to wrong decision-making.

Furthermore, the aspects of qualitative data will be absent from the data collected. From the above example, data stored in the database will not contain, any parametric values related to quality. One instance, maybe the strong correlation index, between strong stereotypes and sales generation. While conveying marketing research, the results of which will be presented in detail in the next sections, we could discern that customer

stereotypes, may not quantify the qualitative aspects of a local brand compared to MAANG companies.

The question now arises, what exactly can be done to measure the qualitative underlying foundation of sales personality? Every data conveys its own story; a concept coined as Data Storytelling. Materialistic aspects of data can be measured, and insights can be generated. When it comes to the personality of customers, pigeonholing to some extent, quality assurance of services, User experience, and information, must be measured separately from quantitative aspects of data. A formula has been developed, which would assist in better comprehending qualitative data parameters.

$$Q \approx pe \pm U$$

Where:

$Q \equiv$  Quality

```
DELIMITER $$
use data;
create index indci_sales
on sales(s_date(100),o_date(100));
select * from sales;
update sales
set shipmentmode='First Class' where cu_id='SC-20725';
lock table sales read;
select * from sales;
unlock tables;
create procedure getDate(IN S_date date)
begin
select s_date from sales
where region='Central';
end $$
select region,sum(profit) as profits from sales group by region;
select distinct cu_id,region,profit
from sales;
DELIMITER $$
```

$P \cong$  Potential of the service

$E \cong$  Expectations

$U =$  Errors In Measurement

This equation can be used, in ways of ranking Viz-a-viz Likert Scale. For instance, asking a customer how he or she will rate his or her product on a scale of 0-5. This though conventional is an effective way of quantifying qualitative data. However, this way of measurement doesn't provide a holistic view of a customer. For instance, the colossal five factors can be used to identify patterns in datasets. A customer's openness may cause the customer to rank higher compared to those who have higher neuroticism. Thus, entrusting your customer crowd to provide accurate data, will be inappropriate. An efficient way, to calculate data can be with the usage of Probability. Consider it this way, while doing your competitor analysis, ABC Company understood the implication of marketing on customers. However, what aspect of personality does that marketing touch? For instance, the probability of profits as per region can help us discern deeper aspects of the region and the opinion of customers.

Table 2: Probability Scores based on Region Wise sales data (2023)

Probability Measures*	Region Wise Areas	Score
$P\alpha$	East	0.9736
$P\beta$	South	0.9816
$P\gamma$	West	0.9220

\*The Probability Calculation was based on sales as well as Customer Ratings.

The data above depicts the probability score of deriving sales region-wise. The data can be perceived to contain lower probability scores in the areas of western parts of Pune. Thus, customer buying behavior tends to be non-favorable. Some variables such as quality assurance, marketing campaigns, and Infitum are assumed to be constant temporally. This data will provide the basis for the codes, used in thematic analysis.

These data can also act as a guiding source, to shortlist the probability of sales from each service mix, of the ABC Company.

Table 3:

Pearson Correlation based on region-wise sales (2023)

Pearson Correlation	Region-Wise	Index
$C_{\alpha}$	East	0.1935
$C_{\beta}$	South	0.1920
$C_{\gamma}$	West	0.1911

Here, the results showcase that West leads to the least amount of correlation between sales and quantity due to the high number of competitors in the market. After finding generalized units of measurement, otherwise known as constants, one can measure patterns. Pattern detection enables a business analyst to comprehend Key Performance indicators. In western regions, ceteris paribus the customers showcased the halo effect, especially in the case of local brands. Suppose the halo effect is measured between the scale -1 to 1. One can make out by the correlation coefficient of west, that the halo effect must be positive. Though it may be difficult to exactly quantify the qualitative aspects, quantitative data can help measure quality parameters.

Navies Bayes Probability for instance can also be used to discern, that due to a lack of marketing campaigns in the West, customers lack brand awareness. Thereby the probability of an increase in sales after raising brand awareness may be predicted to rise. Here one can chalk out qualitative similarities and differences between regions, thus discerning the missing features of data. This will lead to further experimentation with the missing pieces.

1.3 Decision Tree and Random Forest

Case Study: Todorov Judgement Heuristics Experiment

Todorov has found that people judge competence by combining the two dimensions of strength and trustworthiness. Todorov showed his students pictures of men’s faces, sometimes for as little as one-tenth of a second, and asked them to rate the faces on various attributes, including likability and competence. These faces were of Campaigning Politicians.

He then compared the election results to the rating that his students gave. About 70% of the results were in favor of that congressman who was thought to be competent by the students. Thus, the judgment Heuristic comes into play. This striking result was quickly confirmed in national elections in Finland, zoning board elections in England, and various electoral contests in Australia, Groundless, and Mexico.

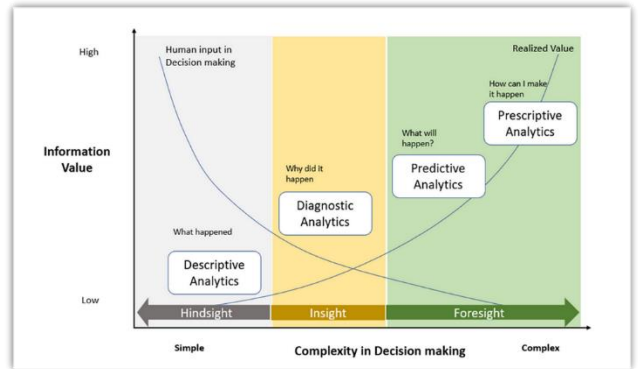
Surprisingly (at least to me), ratings of competence were far more predictive of voting outcomes in Todorov’s study than ratings of likability. However, it cannot predict how well a politician will perform in office. Similarly, when an analyst perceives the data at its face value, he or she can judge how competent in how well a dataset will be in the decision-making process.

Source: Thinking Fast and Slow by Daniel Kahneman

A priori we discussed the aspects of pattern detection in business analytics especially when measuring key metrics. After detecting patterns, and a comprehensive overview of the firm’s key performance indicators we must discern what decision must be made. For instance, consider the above example of sales data. The effective utilization of data determined the correlational index.

That is, before deriving the insights, one must try and define the hindsight either from local servers or from integrated processes. To do so, the firm must have some or the other human inputs. The human inputs decline as the tenure progresses from hindsight to foresight as human intervention in the future will reduce due to uncertainty. The figure below defines the main aspects of the decision-making process:

Figure 5



Analytics maturity model by Milind Desai

During the stage of decision-making, a firm must define standard operating procedures. A business must determine if the answer to a question is yes, what steps will be followed, and if no, what procedure should be undertaken. To do so, a firm defines its decision tree. Suppose the research question is defined as:

“The Western Regional Sales are consistently low in Quarter 3”.

A binary decision tree has a hierarchical structure consisting of a parent node and two child nodes. Each parent node features a split condition that specifies a predictor and the associated cutoff value. [Hong, Lee et al]. The Cutoff value originates from the split condition. This helps regulate impurities in the decision which in turn is termed as information gain. The Node Level Stabilization typically for leaf node which identifies a cutoff value for the most frequent predictors. The decision tree, assuming it consists of only two Boolean values, i.e. Yes or No will look something like this:

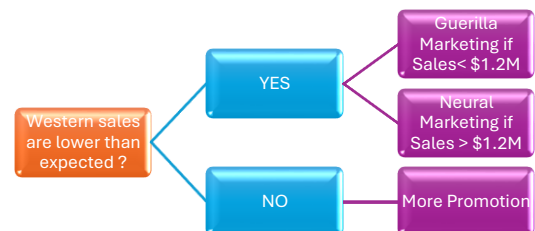


Figure 6

Representation of decision and leaf node in the form of decision tree

The insights driven by the decision tree; define what action will the top-level management take when faced with concerns. A firm realistically must undertake Countless decisions, viz-a-viz Countless calls must be taken in Countless domains. Ergo, just one decision tree will not be enough or unrealistic to go through every decision tree. Taking a final decision a posteriori to the instance highlighted above will involve Countless other decision trees. The combination for the decision tree is called a Random Forest.

Random Forest as defined in [4] is a generic principle of classifier combination that uses  $L$  tree-structured base classifiers  $\{h(X, \Theta_n), N=1,2,3,\dots,L\}$ , where  $X$  denotes the input data and  $\{\Theta_n\}$  is a family of identical and dependent distributed random vectors [International Journal of Computer Science Ahmed, Maqsood]. It involves the input as well as quantities having a direction. For instance, YES will have a positive magnitude while NO in this case also has a Positive Magnitude. A semantically identical tree relates to a Random Forest, which determines the commonalities between the structures of different decision tree models.

Let's discern this model with the help of the Analytics MATURITY model. Suppose a business undertakes a descriptive analysis based on hindsight business data. A Random Forest was generated to discern the anomalies of the businesses earlier on. The  $R_{FRP}$  measured would be subjected to the cutoff value decided by the business. Since the data *a posteriori* involves  $R_{FRP} > \$1$  Million, we can assume the cutoff value as \$1 Million. However not every time will  $R_{FRP}$  be a predictor. For instance, sometimes the expenses can be taken into consideration compared to Sales.

For predictive analysis data derived, another decision tree can be made into a random forest. The random forest can then have the actual  $R_{FRP}$  (Frequently Recurring Predictor) predictions. Suppose the sales are counted to be crossing \$2 Million in next Qtr. The  $R_{FRP}$  here would be \$2 Million. Similarly, the other predictors can be sidelines to focus on the paramount aspects of decisions. This random forest will then be based on prescriptive analysis, which states "How will ABC Company improve the sales probability in Western Region" in the long run? ABC Company may cease Trading in Western Region, till it develops more sales in its current sales territory.

In the next section, we will discern the data-driven insights from the data as well as the research questions, procedures, and methods adopted to spotlight the weak spots of business. We will also undertake statistical Insights of Key metrics along with other datasets utilized by applied business analytics procedures to facilitate business decision-making.

---

### III. RESEARCH METHODS AND PROCEDURES

Though the profits of ABC Company are soaring compared to Q4, in one region in Pune the profits have stagnated. This stagnation was a result of unforeseen events as well as controllable factors.

The research setting is in Pune; It does not assume the demand for the firm to be the same but assumes *ceteris Paribus*. The population was selected, based on the experience of the sample frame in the purchase of Financial Services, I.e. the sample must have used a financial service before. As stated *a priori*, The western region of Pune showcases the lowest probability of sales, though the headquarters of ABC Company is set in the west of Pune.

The research design is descriptive and quantitative also involves some statistical measures of data and visualizations. The datasets used in the study are trifacta. One dataset is the sales data for ABC Company 2023. The second one involves the forecasted and actual revenue and expenses. These datasets comprise the secondary source of data.

The Key Performance indicators are measured using this data. These measures help provide the base, for the primary data collection method, i.e. Survey. The third source of data identifies the information gathered from the participants from west of Pune. This was done to assess the potential demotivating factors for customers, to purchase services offered by ABC Company.

#### 1.1. Sampling Frame

The sampling frame comprises 67 participants, with each participant having an equal chance of selection. N.B the 67 participants were selected from a pool of 80, who voluntarily claimed to use financial services before the survey. Herby some participants were censored based on the above criterion. The possible extraneous variables were also identified. Thus, the genre of sampling used was purposive or judgemental sampling, where the researcher selected the sampling frame with the purpose of identifying key constraints or weak areas of the ABC Company.

#### 1.2. Inventory Design

The survey was administered using online communication means, where the participants volunteered to share the information. The survey was a 10-item scale survey, where 6 items were closed-ended item inventory and the other 4 were open-ended item inventory. The Likert scale, comprised of a 5-point scale, with the response anchors reported to be Strongly Agree and Strongly Disagree. Confidentiality for personal information was maintained and instructions were given in the survey to the participants.

#### 1.3. Procedure of the Research

The research can be divided into three parts. The First consists of discerning and pinpointing the areas of diminishing sales. This would lay the foundation for the second part. Data visualization tools were used to identify potential constraints or areas of improvement. The second step would be identifying gaps in Actual and forecasted data. This enables an analyst to identify vital business positions. If the actual data is better than the forecasted one, it's good weather ahead. If vice-versa, then it's a caution sign for the firm.

Thereafter the third setup consists of administering a survey and collecting data from participants. This data will be used to measure key areas of improvement in a usual context, limiting to ABC Company as a firm. Instead, key areas such as Word of Mouth or Time constraints were identified and measured as per the quantum of participants. For instance, more participants strongly agree with the item, measuring Behaviour in item constraints compared to word of mouth, will assist us in identifying the magnitude of the effect of Time on Potential Customers. The last stage magnifies the participant's viewpoint and thus assists in introducing more stringent measures by the firm.

---

### IV. DATA ANALYSIS AND INTERPRETATION

The current business position will be understood by analysing data from various databases. The emphasis is on quantitative data, involving Countless aspects of economic significant variables. Measures such as Revenue, Expenses, Quantity, and infinitum portray the bigger picture. Suppose a firm, wants to quantify the change of these variables on gross sales. This can be achieved when we know the variables playing a role in affecting sales.

This can only be done by statistical measures and visualizations. A visualization of secondary data assists fills the gap between the actual and the forecasted as discussed earlier. Before the datasets can be used for further processing, a process called ETL takes place. In PowerBI and other visualization software packages, ETL is available to remove Empty Datasets, Arrange the data based on numerical values, or sort the data from A to Z or Z to A. The sales data will be analyzed first to set the foundation for the study.

1.1 Review of Methodology Application

Research Methodology is a vital concern for any research to take into account. Defining the methodology for each dataset, considering this study is a crucial aspect. This will enable the discerning of the genre of data and the data pipeline linkage, to the different genres of datasets acting as nodes. The Methodology this study adopts is twofold. The sales data, using exploratory analytical research methodology, explore different regions and service industry sales. The Forecast data uses descriptive-analytical research to describe the flaws in data. The exploration is limited to the sales and the Profit Percentage, categorized by each service offered by ABC Company as well as the Region-wide distribution. The application of Methodology adopted for the databases is divided as per datasets:

A) Sales Dataset

Exploration, of sales differentiation between services offered by ABC Company as well as Region Wise sales distribution is conducted. The explored datasets are then used for further analytical processing, to interpret the data, i.e. the Business Shortcomings. These analytical processes also aid in highlighting the errors. The distribution below marks the advent of the application of the methodology adopted:

Table 4 Sales Distribution as per services offered

Service Offered	Sales (in K)	Profit Percentage
FundsPay	246.1 K	2.361 %
FundsResearch	271.7 K	2.363 %
FundsTech	215.3 K	2.364 %

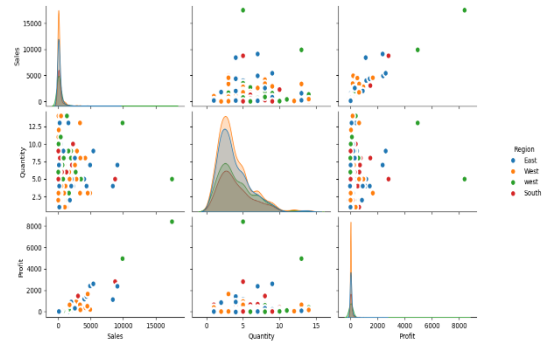
The profit percentage is analysed based on every service offered and is found to be almost equal. Thus, we can interpret from this data that, ABC Company should Increase asset allocation to FundsPay, to improve the margin by 0.003 %. Additionally, ABC Company can increase the Forecasted Targets for FundsPay. Based on Region, the distribution of data will be as follows:

Table 5 Region-Wise Sales Distribution

Service Offered	Sales (in K)	Profit Percentage
West	238.5 K	2.361 %
East	248.2 K	2.363 %
South	246.4 K	2.364 %

Figure 7

Graphs matrix to measure regional sales with every known variable



Similar Results can be found here. The sale distribution from East, west, and south. Thereby the interpretation would be to raise capital expenditure or budget on promotion in Western regions. However, as an analyst, you must discern the factors underlying the data. For instance, the profit percentage is lower in the west region of Pune compared to the east and south. Thereby the percentage was caused to be lower than the other two regions. Consider an example:

$$Z = \frac{X}{Y}$$

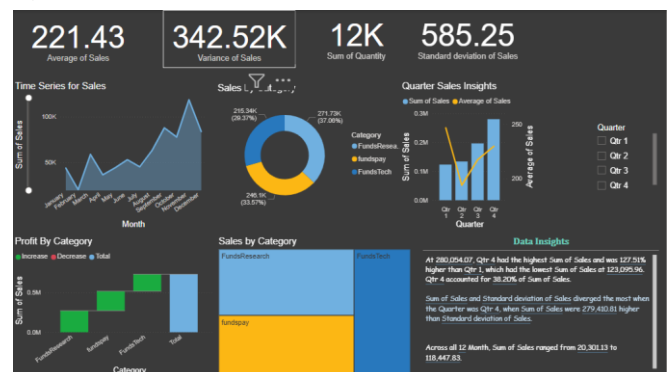
When Y increases, X will decrease.

$$Z = \frac{X}{Y + 10}$$

Other statistical measures will be affected by more selling orders for Western Regions. Another aspect that comes into play is the error measurement. (U) will be involved while measuring the data or data collection. This is nothing but measurement errors. Thus, more the number of calculated participants, the lesser will be the error as an increase in sample size will lead to a reduction in error. This is called the Principle of Inertia of Large Numbers.

A priori, discussed probability measurement or correlational measurements can help describe the anomalies in measurements and bring more goodness to results. Thus, Western regions of Pune have less probability of sales (0.9220). This implies that there exists a problem in Marketing in western regions of Pune, though there may be more underlying assumptions held by the customers of FinServ Industries. Data is as good to use as its perspective. Analysts must analyse the data from different viewpoints. The following represents the exact analysis of sales data, in the form of a dashboard:

Figure 8



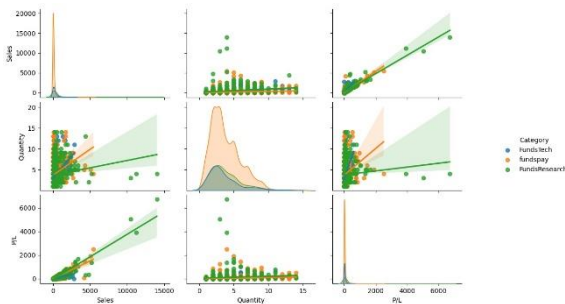
Analysis of sales data

The interpretation of the datum utilized in the dashboard is as follows:

- I. The breakout point for the firm is Qtr3. This Qtr. See exponential growth compared to previous quarters. This is so, because of seasonal variation. For instance, Festival sees investments in Gold or other precious metals.
- II. The market share of Fundtech is lower than the other services. The reason might be the service differentiator and the presence of competitors in the same field. The newly launched Fundtech, might not be promoted enough to draw sales since its inception.
- III. The average sales Dips in Qtr2 to rise again in Qtr3. The reason why the average may have dipped would be the beginning of the financial year. Usually in March, the demand for auditing services increases, especially in the case of ITRs. In April the demand will fall sharply due to the beginning of a new financial year.

After examining sales data, we understood the aspects of datasets and interpretations of the data. However, to discern the business position better, and to promote accurate prediction, the forecast data and the actual data of 2023 are put into effectuation. The next section deals with The Gaps between the actual or real data vs the perceived or forecasted datum.

Figure 9



Graphs matrix to measure services-wise sales with every known variable<sup>3</sup>

B) Forecast Vs Actual Data

This dataset focuses on describing flaws in assumptions between forecasted data and the actual dataset. Furthermore, the data is subjected to statistical functions to portray the anomalies or errors in forecasted information. One example of forecasting is a neural network. Firms try to develop a system that mimics a human mind, consisting of complex computational webs. One paramount criterion of the neural network is Pattern Detection.

Therefore, the study of forecast and actual data tries to detect the magnitude of changes between Actual and Forecast Data. Below, is the descriptive statistics for Actual Revenue 2023 by ABC Company, obtained using the data analytics function in Microsoft Excel:

Table 6

<sup>3</sup> Heteroscedasticity can be visually inspected under these visualizations, based on scattering detected.

Actual Revenue 2023 Descriptive Statistics

Measures	Amounts
Mean (X)	86.844
Median	62
Mode	52
Standard Deviation	66.183
Kurtosis	2.509
Skewness	1.757
Range	364
Confidence Level (95%)	2.163

Actual Revenue

The distribution involves many statistical measures ranging from Kurtosis to median. Though the Median and other datum can be compared, the same is not true for standard deviation. To facilitate comparison, the Correlation of Variance is calculated:

$$CV = \frac{\sigma}{x}$$

Where  $\sigma$  = standard deviation

X= Mean

The CV is valued at 0.76. This means that there is a lesser Standard deviation as  $\sigma < 1$ . Furthermore, the Kurtosis of the measure can also be assessed. Since  $K < 3$ , the distribution shows a platykurtic distribution of outliers or deviates. The interpretation would represent consistency in the Actual data. Thereby representing more equalness or closeness to the Mean.

In comparison, the data of forecasted revenue typically descriptive statistics is also mentioned below:

Table 7

Forecasted Revenue 2023 Descriptive Statistics

Measures	Amounts
Mean	77.866
Median	50
Mode	40
Standard Deviation	66.256
Kurtosis	4.986
Skewness	2.103
Range	450
Confidence Level (95%)	2.165

Forecast Revenue



The distribution involves numerous statistical measures ranging from Kurtosis to median. Though the Median and other datum can be compared, the same is not true for standard deviation. To facilitate comparison, the Correlation of Variance is calculated:

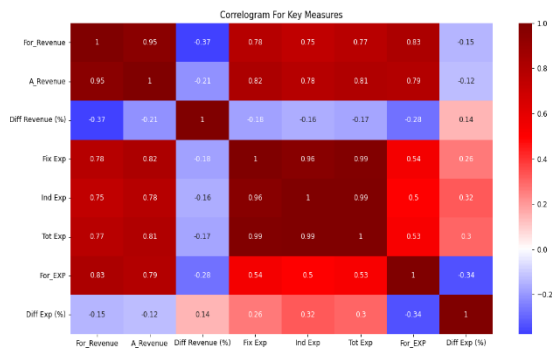
$$CV = \frac{\sigma}{x}$$

Where  $\sigma$  = standard deviation  
 $X$  = Mean

The CV is valued at 0.85. This means that there is a lesser Standard deviation as  $\sigma < 1$ . Furthermore, the Kurtosis of the measure can also be assessed. Since  $K > 3$ , the distribution shows a leptokurtic distribution of outliers or deviates. The interpretation would misrepresent consistency in the Forecasted data. Thereby representing more deviation from the Mean. Generally, there is a human tendency of overidentification of the forecasts. In a model of  $M$  simultaneous equations, for an equation to be identified, it must exclude at least  $M - 1$  variables (endogenous as well as predetermined) appearing in the model. If it excludes exactly  $M - 1$  variables, the equation is just identified. If it excludes more than  $M - 1$  variables, it is overidentified. [Gujrati, Porter]. Thereby only the most proximate variable must be identified.

One way of undermining the variables would be correlating them with each other. For instance, the highest correlation must be taken into account ( $> 1$ ) to forecast or predict values. ARIMA is a common example of the above. Below is a correlogram that displays correlations of each variable in the context of another variable

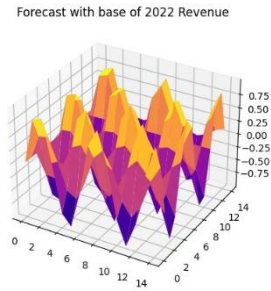
Figure 10



Correlogram for actual and forecasted data

Another example, to determine the forecast for 2023 Revenue, will be with the base variable being 2022 Revenue similar to exponential smoothing. The figure depicts the forecasted graph.

Figure 11



Forecast of revenue with 2022 revenue as z constraint

Thus, key interpretations of the forecasted datasets and the actual datasets, also involving the earlier sales data are:

- I. The correlational index between For\_Exp (Forecasted Expenses) and For\_Revenue (Forecasting Revenue) is 0.83. This implies a positive and strong correlation between For\_Revenue and For\_Exp as predicted in 2022. This denotes that, if a firm assumes or predicts to have more revenue, the firm will increase its expenditure. One key error can be, the lack of a backup plan. If the firm fails to reach the target, the firm will be in a tough situation as the expenses were done as forecasted. This is serious as for a customer, one bad experience is enough to increase the firm's Customer Acquisition Cost (CAC).
- II. The correlational Index between A\_Revenue (Actual Revenue) and For\_Revenue (Forecasted Revenue) is 0.95. This implies that actual revenue should be taken into account to forecast operations. For instance, a firm should account for a major share of forecasting, to Actual Revenue compared to their variables. A strong demotivator for a potential customer is the market share, a firm enjoys. Autocorrelation can be used as a measure for more accuracy and predictability.

Risk assessment can be done on distribution kurtosis. Platykurtic consists of flatter tails, which means fewer outliers. This means, that Actual Data has a lesser risk to invest in. Leptokurtic consists of fatter tails, comprising more outliers. What this implies is that the forecasted data implies more risk compared to actual ones because it has not happened. Though this does not imply the risk, portrayed by the customer, he or she might perceive a firm as risky if they face loss aversion. For instance, a customer at the time of filing his ITR, will tend to avoid risk and subscribe to well-known brands compared to Local ones.

The forecasted data sets the baseline, to answer the research questions. Correlation of variables, Kurtosis, and Coefficient of variance estimates are necessary to discern the business position and the flaws in data collection. The sales data a priori, is an actual dataset. If the forecast values are not met, it means two things. First, either the firm is not utilizing its full capacity. This can be due to a lack of capital funds or shareholder disagreements. Another aspect, of not meeting forecasted targets, could be a lack of public interest in the firm. The lack of interest could be because of the presence of extraneous variables such as Time, Experience etcetera. These variables are uncontrollable. However, this variable can be used to identify the magnitude of the effect of customer behavior. How exactly will Urgency affect the customer? To what magnitude will urgency negatively reinforce a customer to lose interest in investing in ABC Company Services? These questions will be answered in the next section.

## 1.2 Review of Research Questions

When you are in a state of cognitive ease, you are probably in a good mood, like what you see, believe what you hear, trust your intuition, and feel that the current situation is comfortably familiar. You are also likely to be relatively casual and superficial in your thinking [Kahneman]. Continuing the above example, A customer in an easy setting will purchase services from well-known brands. The effect of urgency, or lack of interest will be prevalent in this setting. Suppose a customer has two pathways to take. He wants his ITR Filed, and he has two options, either to purchase from a tried and tested brand in the market, though he may not have tried the service beforehand. Another option is the service from a local brand. Chances are that the customer in an easy setting will purchase service from a tried and tested product.

When you feel strained, you are more likely to be vigilant and suspicious, invest more effort in what you are doing, feel less comfortable, and make fewer errors, but you also are less intuitive and less creative than usual. [Kahneman]. The instance highlighted A priori, the chance of customer purchasing service from the Local Brand increases. Thus, the best chance at generating revenue for a firm is to target a customer when the requirement for the service becomes threatening. For instance, targeting the customer in March, when the need to fill the ITR arises will be beneficial.

The research questions, this study tries to answer are threefold. These Questions which sets the background of the main question are:

- I. What Regions or Services have the lowest profit-generating behavior?
- II. How can a business effectuate forecasting? What variable can come into play?

The main research question is:

- I. What extraneous variables play a part in selling?

Overall, I and II were answered by the section above. In this section, we will discern Question III, i.e., the market components can be targeted for better results. The components are as follows:

1. Experience
2. Society
3. Opportunity Perception
4. Word of Mouth
5. Urgency or Time Constraints
6. Purchasing Magnitude<sup>4</sup>

## 1.3 Findings of the study

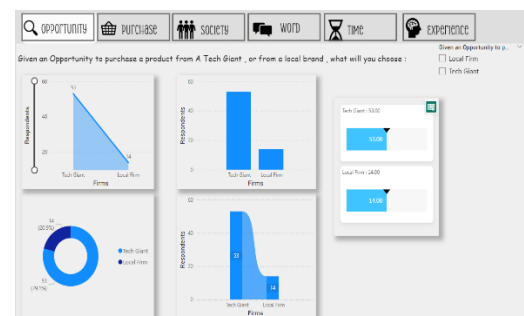
The study aimed to undermine the magnitude of factors affecting the customers which in turn would affect the sales performance of the firm. Overall, the proportion of participants lies in between strongly agree and agree range values. Therefore, if we were to quantify the values allocated to them, the magnitude lies between 4 and 5. Thereby the extraneous components, i.e. Experience, Society, Opportunity Perception, Word of Mouth, Urgency or Time Constraints, and Purchasing Magnitude can be used as extraneous variables to develop a sales equation. The details for each item measuring the components' magnitude were visualized using data visualization. The closed data inventory acts as evidence for open-ended questions. For example, one of the items asked was

“Would you trust a Fortune500 company, after a bad customer service experience (assuming you have built a relation for five years)”

This can act as evidence to prove Experience Metrics as measured *a posteriori* as many agreed to trust the Fortune500 Firm irrespective of a bad customer service experience. The term opportunity perception refers to as to what extent will a customer chose a tech giant compared to a local brand, especially in FinTech Settings. The Items asked in the inventory were:

“Given an Opportunity to purchase a product from A Tech Giant, or a local brand, what will you choose”

Figure 12



Opportunity metrics

Here, in an easy setting, 53 participants chose Tech Giants while 14 participants chose Local Firm. Thereby, we can analyse the profile of the customer, viz the non-risk-taking behaviour of customers.

Next, the Purchasing behaviour of the customer is defined as the extent to which a customer will go to purchase a financial product or service. This item was also censored while censoring all those participants who selected Disagree or Strongly Disagree.

“You are a person who frequently purchases financial products”

The participants choosing Strongly Agree and Agree are nearly equal. This shows us the seriousness of the participants towards financial services. More number of Agree shows us, that the customers in the region are not too dependent on Financial Services. Another component is one of the most crucial aspects of customer behaviour. Society leads customers to purchase in-demand products and filter out the Local Brands considering them to be of lesser quality.

“Societal Pressure affects you to purchase products with more brand value”

<sup>4</sup> NB: The 80 participants were asked the same items. Disagree and Strongly Disagree were censored.

More numbers, of Agree and Strongly Agree depict that customers will be affected by what society thinks of the brand. A local brand may not be perceived as well as a Giant one, though sometimes a Local brand may provide quality service at a lower price. The next component is words. The word-of-mouth phenomenon is vital when discerning customer behaviour, and is directly proportional to sales generation.

“You are demotivated to purchase a product if you are unaware of people who have tried the product before.

Figure 13



Word of mouth metrics

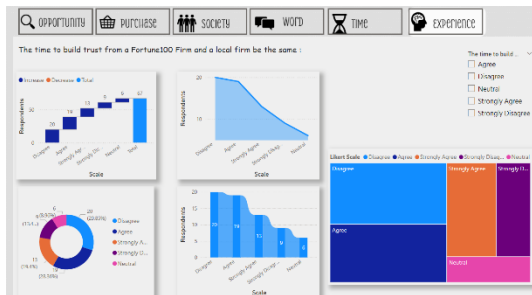
As good as word of mouth may occur, similarly nonword of mouth will occur. If a customer doesn't know anyone who has tried and tested the product before, he or she will not purchase it. As seen in the leptokurtic behaviour of forecasted data, the risk will be multiplied with additional risk perception. The component of urgency or time constraint signifies a higher magnitude of effectiveness. The more urgent the task, the lesser intuition a customer will have.

“In a situation of urgency, you search for a financial service online. Which website will you click”

The first few ones have higher participant engagement. This means, that a customer during an emergency will purchase the first brand that he encounters. The last but the major component is Experience. A customer builds trust either by brand value or by experience. If a customer finds consistency in service, especially a local firm, the customer will gradually be a permanent customer.

“The time to build trust between a Fortune100 Firm and a local firm be the same”

Figure 14



<sup>5</sup> Inversely Proportional was calculated based on the maximum responses recorded (Here Disagree was more)

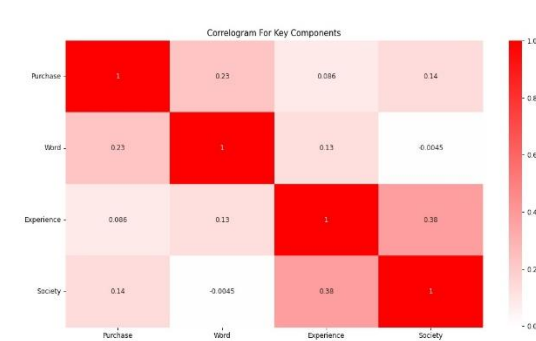
Experience metrics.

Major trends showcase the disagreement of the participants towards the time built to generate trust. An equal amount of agreement showcases the perception of the customer. There is a high chance, that the participants agreeing to the item, have lost the perception of local brands associated with bad quality or experience.

Thus, each component has an equal role to play in promotion and risk mitigation.

The extraneous elements' magnitude is measured using the visualization generated. This enabled the application of real-time elements when forecasting the datasets. Thus, another way to discern the variables in the dataset would be extraneous correlations, among major variables. To measure these components properly, the scales (1 to 5) assigned to them by participants can be converted into correlational ranges especially Society, Purchase, Word, and Experience due to the involvement of 5 scale measures. The correlogram for these measures:

Figure 15



Correlogram for survey metrics

The highest correlation (< 1) is 0.38 between Society and Experience and vice versa. This tells us the key areas to be taken into account from the survey. Assuming we keep these two components as variables of the sales equation. Let A represent Society and B represent Experience. Thereby the sales proportion will be:

$$\Delta \text{Gross Sales} \propto \Delta A$$

$$\Delta \text{Gross Sales} \propto 1/\Delta B^5$$

Therefore

$$\Delta \text{Gross Sales} = \Delta A / \Delta B \dots \dots \dots \text{Eq 1}$$

The variables Word (c) and Purchase (d) can also be included in another equation. That is:

$$\Delta \text{Gross Sales} \propto \Delta c$$

$$\Delta \text{Gross Sales} \propto \Delta d$$

Therefore

$$\Delta \text{Gross Sales} = \Delta c + \Delta d \dots \dots \dots \text{Eq 2}$$

$$\Delta \text{Gross Sales} = \Delta A / \Delta d \dots \dots \dots \text{Eq 3}$$

The variables from the predictive model were Forecasted Revenue (X) as well as Actual Revenue (Y). Thereby another equation would be represented as:

$$\Delta \text{Gross Sales} \propto \Delta X$$

$$\Delta \text{Gross Sales} \propto \Delta Y$$

Therefore

$$\Delta \text{Gross Sales} = \Delta X + \Delta Y \dots\dots\dots \text{Eq 4}$$

$$\Delta \text{Gross Sales} = \Delta c + \Delta Y + U \dots\dots\dots \text{Eq 5}$$

Thereby, this equation follows the M-1 Rule, where M=5. Thus, Just Identification was identified. This equation can be used, to analyze the prediction as well as the Gross sales in short-term tenure. For instance, suppose ABC Company wants to predict the sales in the next quarter. The firm just must express the variables in terms of the magnitude of change to receive the gross sales up to a decent accuracy. The magnitude or probability of the event can be measured simultaneously to reach the result.

As seen in Random Forest Analysis, the Boolean values of YES or NO may have positive or negative magnitude. Suppose the top management wants to consider increasing the forecasted revenue, a controllable variable presently. To do so, a firm will have to incorporate every variable in the equations while deciding what leaf node to pick. For example, after selecting YES as a leaf node, the other nodes such as societal factors must be considered as a change in society will ultimately lead to a change in Forecast Revenue.

## REFERENCES

- JEONGEON LEE, & JUNG-SIK HONG. (n.d.). *Assessing Decision Tree Stability: A Comprehensive Method for Generating a Stable Decision Tree* [Review of *Assessing Decision Tree Stability: A Comprehensive Method for Generating a Stable Decision Tree*].
- Jehad Ali, & Rehanullah Khan. (2012, September). *Random Forests and Decision Trees* [Review of *Random Forests and Decision Trees*].
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- Cosma Rohilla Shalizi. (n.d.). *Advanced Data Analysis from an Elementary Point of View* [Review of *Advanced Data Analysis from an Elementary Point of View*].
- Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics*. McGraw-Hill/Irwin.
- Abideen Mayowa Abdul-Yekeen. (n.d.). *Business Analytics and Decision Science: A Review of Techniques in Strategic Business Decision-Making* [Review of *Business Analytics and Decision Science: A Review of Techniques in Strategic Business Decision-Making*].
- Ulmer, R. R., Sellnow, T. L., & Seeger, M. W. (2019). *Effective crisis communication: Moving from crisis to opportunity*. Sage Publications, Inc.
- Arun Kumar, Robert McCann, Jeffrey Naughton, & Jignesh M. Pate. (n.d.). *Model Selection Management Systems: The Next Frontier of Advanced Analytics* [Review of *Model Selection Management Systems: The Next Frontier of Advanced Analyt*