FINAL SUMMER PROJECT REPORT

On

Data Analytics with Python

Submitted by:

Name: Anshuman Singh Deora

Roll No: 201107

Section: A

Batch: MBA–FT (2020-2022)

Institute of Management, Nirma University

Date of Submission: 12-07-2021

TITLE DETAILS

Author of Report	Anshuman Singh Deora
Company Name	Gateway Technolabs Pvt. Ltd
Address of the Company	B-81 Corporate House, S.G. Highway, Bodakdev, Ahmedabad, Gujarat 380054
Project Title	Data Analytics with Python (Finance and Investment Strategies in Indian Market)
Date of Report	12.07.2021
Mentor	Mr. Niraj Gemawat
Purpose of Report	Partial Fulfilment for the award of Degree of Master of Business Administration for the Batch 2020-2022
Presented to:	Institute of Management, Nirma University
Faculty Mentor :	Prof. Harismita Trivedi

ACKNOWLDEMENT

I had internship opportunity with Gateway Group. Despite the fact that it was a virtual internship, it provided an excellent opportunity for learning and professional development. As a result, I consider myself really fortunate to have been given the opportunity to be a part of it. I also had the opportunity to meet a lot of new people and professionals that guided me during my internship.

I would like to thank **Mr. Niraj Gemawat** for consulting, supporting and guiding me throughout with patience and immense knowledge. I would not be able to finish this project without his guidance and encouragement. I have gained lateral and in-depth knowledge of the assigned topics due to this Summer Internship. Also, I would also like to extend my thanks to the Institute for giving me such a wonderful opportunity to work with such a great organization.

Lastly, I would like to express my sincere thanks to my faculty mentor **Prof. Harismita Trivedi**, for providing me with the necessary guidance regarding the important things before the start of my internship as well during the course of internship. She was a thorough support and has always been encouraging and appreciative of the work.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

During the Internship at Gateway Group which started on 05-05-2021 which aimed to give knowledge about finance and investment strategies in Indian market. It teaches us about the current market scenario by the help of some Data Analytics Technique with python. The Data analysis teaches us few financial and technical skills that are required for filtering data, data collection, data extraction, data visualization through charts and bars graphs.

On the first day we had a proper formal training and briefing session where explanation of the company and working of the company was explained. I was already allotted with mentors and later after induction they provided me access to portal like Anaconda on which I have learnt and performed programming on python. As a part of internship I was asked to learn few of the strategy and their regular functioning in the day to day scenario and after that I was supposed to learn about the execution of algorithms and on the daily basis of the status. I use to execute those algorithms on the platform like JupyterLab on Python. The whole idea is to have the perfect strategy in place post which they can takeit to scale through building a technology product around the same. It will help them to generate and track the data at any point of time. This experience was challenging as well as full of opportunities.

Part A

PROFILE OF THE ORGANISATION

About the Company:

The Gateway Corp. is the Gateway Group's primary holding company and promoter of independent operational firms. The company headquartered are situated at India(Ahmedabad) and The Netherlands.

The Gateway Corporation is a for-profit corporation that operates independently under the direction and supervision of its Board of Directors and Shareholders. It is made up of businesses that specialise in disruptive innovation, industry-oriented technology consulting, software platforms, and IT services.

Gateway Group was founded with a single goal in mind: to help our customers succeed in the markets they serve. We've been expanding our wings over the globe with different customer winning tales thanks to the pooled wisdom of the founders and a multinational global workforce.

The entrepreneurial genome of Gateway Group is its fundamental differentiator. Over the last two decades, our clients throughout the world have experienced the foundation ideals of honesty, ethics, and devotion that have driven the organisation to surpass obstacles. The group, in all of its activities and brands, is able to flourish because of an egalitarian landscape of skills, expertise, and track record, as well as a deep belief in our value system.

Using the most in-demand, fast-growing relevant technology to deliver entire solutions in the most crucial business areas. Create a professionally engaging and cheerful atmosphere for our team where professionals thrive alongside the growth of our company by consistently delivering creative, comprehensive, efficient, and cost-effective business solutions for the worldwide market through IT.

Everything we do revolve around software engineering. We use engineering methods to make the world more better place in technology field. We are a Digital Enterprise guiding brands into the future and assisting them in accelerating their Digital Transformation. We create Unique Digital Solutions to optimise Customer's Business Models, resulting in Better Performance and Greater Customer Delight, with our Consulting expertise, On-Ground presence in 16 countries, and Domain Knowledge gathered over two decades. Our clients refer to us as the "One-Stop Solution Provider" since we strive for excellence and offer:

- Commitment towards customer success
- Combination of technology and domain expertise
- Adoption and continuous improvisation of our software development practices
- Consistent upgradation and training of our team by experts to understand the criticalities and reduce the cultural gap.

The company functions on its own, guided and supervised by its board of directors and stockholders. It includes joint ventures and publicly traded companies that are focused on disruptive innovation and research, industry-oriented technology consultancy, pioneering business-technology platforms, and software services.

Vision

Create a professionally engaging and cheerful atmosphere for our team where professionals thrive alongside our company's growth by consistently delivering creative, comprehensive, efficient, and cost-effective Business solutions through IT for the worldwide market.

Mission

Using the most in-demand, fast-growing relevant technology to deliver entire solutions in the most crucial business areas.

Company's Services:-

1. **Consulting Services: -** The company follows a collaborative design-led consulting

approach to help enterprises envision and implement Digital Transformation. Business services can be divided into the following sections:

- Digital Marketing strategy
- Automative Aftermarket
- Alvarium
- Remote Engineering
- Digital Logistic

2. **Digital Services: -** Enhancing Digital Transformation experiences by combining technology and creativity in a seamless manner. Business services can be divided into the following sections:

   - Enterprise Application Development
   - Product Engineering
   - Cloud Transformation
   - Quality Engineering
   - Cyber Security Services

3. **Professional Services: -** We are a global leader in emphasising the human element of any company's strategy and assisting them in making better business decisions.

   - TecBridge

4. **Future Engineering :-** With futuristic technologies and solutions, organisations can react to quick changes in the business paradigm.

   - 5G
   - Blockchain
   - Internet of Everything
   - Robotic Process Automation
   - Cognitive AI

Company's Products:-

1) **Gateway Digital: -** The gateway digital is transforming the legacies into profitable and remarkable future. It will transform the technology, business process transformation, strategy and digital consulting.
   - Smart Mobility
   - Augmented Reality
   - Extended Reality
   - Machine Learning
   - Block chain
   - Robotic Process Automation
   - Internet of Things
   - Cloud Transformation
   - Strategic Intelligence
   - Cyber Security

2) **DILX: -** The logistic chain real time control with customized services from assessment to ROI. It reduced total cost of innovation with the help of offshore development centers and onsite presence with world class technology solution based on open source and cloud.
   - Control Tower
   - Customer Interaction Platform
   - Market Place
   - Transport Management System
   - Warehouse Management System
   - Freight Management System

3) AutoDap :- AutoDAP is a DaaS subscription service for the automotive aftermarket

that is powered by an API. We help businesses adapt faster by providing accurate and effective data that leads to increased process efficiency, reduced risk, and increased profitability, thanks to our broad experience and unique insight of the European automotive sector.

- License Plate to Basic Vehicle Information (LPBVI)
- Precise Vehicle Identification and Build Sheet (VINVIS)
- WLTP (VINVIS)
- OE Parts (VISOEP)
- Repair & Maintenance Information (VISRMI)

4) **AutoFacets :-** Known as a partner with a difference in the automotive business, and driven by a passion for automobiles. It's in creation of new Digital Service Models, providing vital Tactical & Technological edge.
   - Automotive
   - IOT
   - Big Data
   - Cloud Enablement

5) **G Secure Labs :-** It use to protect your business data from known and unknown cyber threats and the segments they serve are:-
   - Automotive
   - Pharma
   - Energy Sector
   - Financial Services
   - Healthcare
   - Legal Services
   - Transportation

- Info. Technologies
- Telecom

6) **TecBridge :-** In the most culturally complicated business circumstances, the ability to identify and place the right talent is essential.
   - IT Staffing
   - IT Outsourcing
7) **Alvarium** :- Collaborative workshop engagement provides practical insights to stakeholders about the organization's growth path and reduces risks. It identifies and aids in the correction of core issues. It helps in achieving the company goals and objective with a take of digital transformation.

8) **Business Analytics:-** At every 5 minutes, actionable analytics enable global business insights.
   - Industry Analytics
   - Service Analytics
   - Function Analytics

9) **Cloud Services: -** The services offered in our products is one of Europe's largest HRM services that saw a 25% reduction in migration cycles and a 40% reduction in capex.
   - Iaas
   - Saas
   - Paas
10) **Application Engineering: -** It delivers the Fastest competition in delivering game-changing, cost-effective applications in the market.
   - Application Development
   - Legacy Migration
   - Enterprise Developmen

Part B

PROJECT WORK

Purpose of the project:

As an intern, Data analytics can help businesses make better decisions and reduce financial losses. Predictive analytics can predict what will happen as a result of business changes, while prescriptive analytics can recommend how the business should respond to these changes.

Aside from that, my assignment is to learn about the data preparation, data modeling, data evaluation and data visualizations to analyse the outcomes of various data and information and make it useful for the company and business.

Objective of the project:-

The objectives/expectations of the project include:

- Retrieve data from one or more sources and prepare the data so it is ready for numerical and categorical analysis.
- Data exploration is to find interesting trends or relationships in the data that could bring value to a business.
- Building a foundation to read and execute the probability and statistics to analyse and understand the data.
- To create plots and charts to help communicate your data and findings visually.

The activities associated with the project include:

- Exploring the data
- Data Cleaning and Preparation
- Building formal statistical knowledge.
- Interpreting the results.
- Creating data visualization
- Communicating the results.
- Maintaining good contact with the potential leads of the company.

Introduction:-

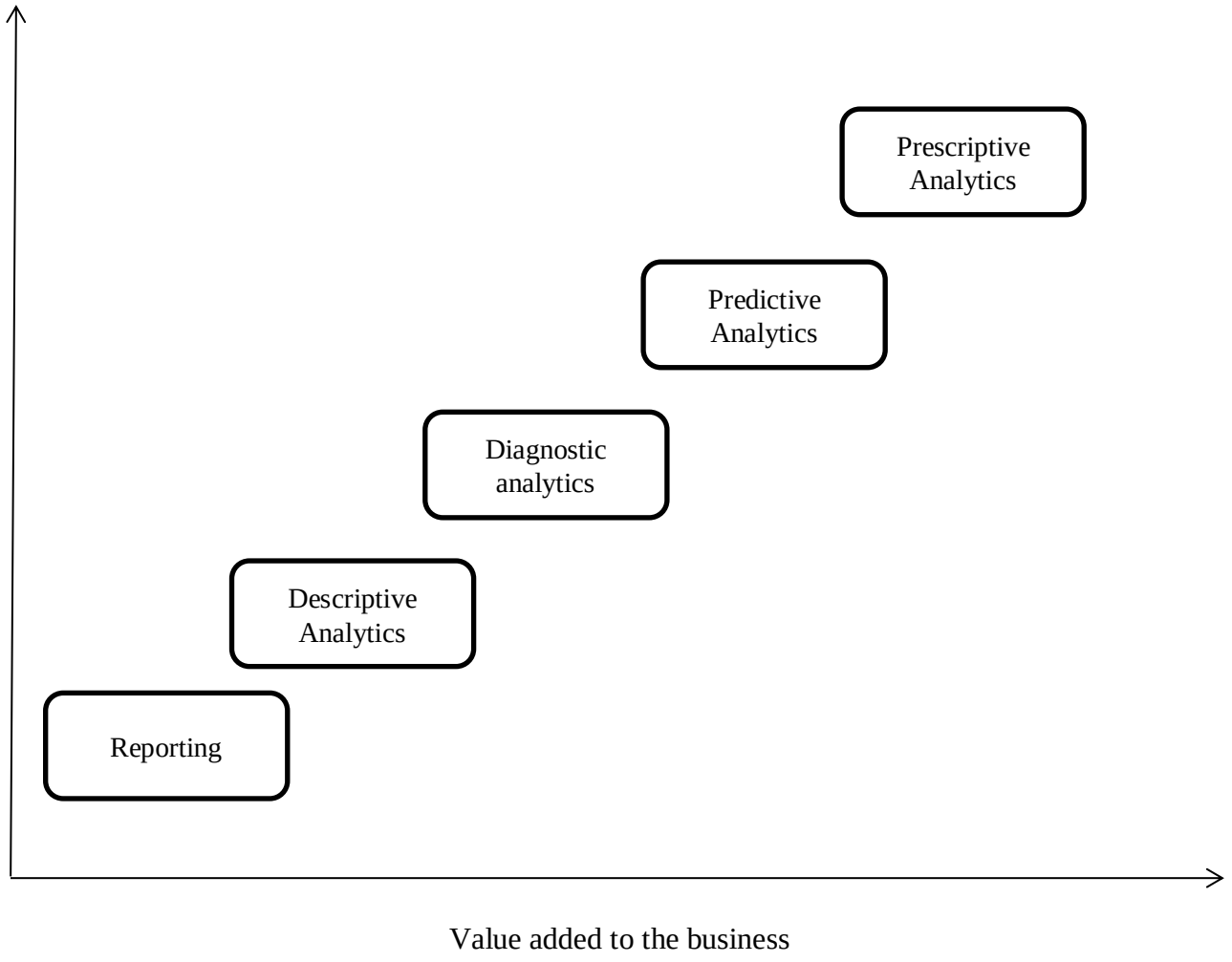
Data Analytics

Cleaning, analysing, interpreting, and visualising data to generate important insights that drive smarter and more effective business decisions is known as data analysis. Data analysis tools extract usable information from corporate data and aid in the data analysis process.

Data science is the study of extracting insights from data using scientific methods.

These technologies are used as for example the large sets of image data from several items in the same category must be processed in order for a computer to recognize an image. E.g. face recognition.

How data analysis helps in business analytics?



Types of Data Analytics

a) **<u>Reporting / Management Information System</u> :-** To keep track on what's going on in the company.

b) **<u>Descriptive Analytics :</u>-** Descriptive analytics aids in the investigation of events. To describe outcomes to stakeholders, these strategies synthesise big datasets. These tactics can help track successes and failures by establishing key performance indicators (KPIs). In many businesses, metrics like return on investment (ROI) are used. To track success in certain industries, specialised metrics are devised. This procedure necessitates the gathering of relevant data, data processing, data analysis, and data visualization. This procedure provides crucial information about previous performance.

c) **<u>Diagonostic Analytics :</u>-** Diagnostic analytics can assist you figure out why things happened the way they did. These strategies are used in conjunction with more basic descriptive analytics. They take the results of descriptive analytics and delve further to discover the root of the problem. The performance indicators are looked at further to see why they have improved or deteriorated. This usually happens in three stages:

- Recognize any discrepancies in the data. These could be unanticipated changes in a statistic or a market.
- Data pertaining to these anomalies is gathered.
- To uncover links and trends that explain these abnormalities, statistical approaches are applied.

d) **Predictive Analytics :-** Predictive analytics aids in determining what will occur in the future. These methods make use of historical data to uncover patterns and decide if they are likely to repeat again. Predictive analytical tools use a number of statistical and machine learning techniques, such as neural networks, decision trees, and regression analysis, to predict what will happen in the future.

e) **Prescriptive Analytics:-** Prescriptive analytics assists in determining what should be done. Data-driven decisions can be made utilising predictive analytics insights. In the face of uncertainty, this enables firms to make educated judgments. Machine learning strategies are used in predictive analytics techniques to detect trends in massive datasets.

The chance of various outcomes can be determined by evaluating past decisions and events.

A data analyst's works include dealing with data throughout the data analysis process. This entails a variety of data manipulation techniques. The importance and balance of these procedures are determined on the data used and the analysis' purpose.

For many data analytics activities, data mining is a necessary step. This entails obtaining information from unstructured data sources. Written text, massive complicated databases, and raw sensor data are examples of these. The extraction, transformation, and loading of data are the most important processes in this process. These steps transform raw data into a format that can be used and managed. This is how data is prepared for storage and analysis. The most time-consuming phase in the data analysis process is usually data mining.

Another important part of a data analyst's work is data management or data warehousing. Data warehousing entails creating and implementing databases that make data mining results accessible. In most cases, this stage entails building and managing SQL databases. Non-relational and NoSQL databases are also gaining popularity.

Analysts can derive insights from data through statistical analysis. Data is analysed using both statistics and machine learning approaches. Big data is employed in the development of statistical models that show data trends. These models can then be used to create predictions and inform decision-making using new data. This procedure requires statistical programming languages like R or Python (with pandas). Advanced analysis is also possible because to open source libraries and packages like JpyterLab.

Data presentation is the last step in most data analytics procedures. This step provides for the dissemination of information to stakeholders. The most crucial tool in data presentation is often data visualisation. Compelling visualisations may assist executives and managers comprehend the significance of these findings by telling the storey in the data.

Python Introduction

Python is a high-level, interpreted programming language that may be used for a variety of tasks. It has efficient high-level data structures and an object-oriented programming technique that is simple but effective. Python's syntax and dynamic typing, as well as its interpreted nature, make it an excellent language for scripting and quick application development across a wide range of platforms.

Python for Data Analysis:

Why Python???

1. Python is an open source language.
2. The Programming syntax is as easy and simple as English.
3. It is widely used language and collaborative developer community.
4. It includes many extensive Packages.

- UNDERSTANDING OPERATORS:

  Theory of operators: - Operators are symbolic representation of Mathematical tasks.

- VARIABLES AND DATATYPES:

  Variables are named bounded to objects.

  CONDITIONA STATEMENTS: If-else statements (Single condition)
  If- elif- else statements (Multiple Condition)

5. LOOPING CONSTRUCTS: For loop

6. FUNCTIONS:
   The Functions are small pieces of programming code that can be reused. Created to solve a specific issue. There are two sorts of functions: built-in and user-defined.
   Python does not allow you to reuse functions.

- DATA STRUCTURES:

Two types of Data structures:

Lists: A list is a data structure that is ordered and has elements separated by commas and enclosed in square brackets.{}

Dictionary: This is usually an unordered data structure with elements that are separated by comma and stored as key in the program like value pair, closed with curly braces {}.

Outliers

A value that is outside the data's range is referred to as an outlier.

Eg- 9300 instead of 93. Reasons of Outliers

- Typos-During the data collection. Eg-adding zero by mistake.
- Measurement Error- Data outliers due to a poor measurement operator.
- Intentional Error- Errors that are purposefully introduced. For example, pretending to have drunk less alcohol than was actually consumed.
- Legit Outlier- These are values that are not errors, but are present in data for legitimate reasons. For example, a CEO's salary may be significantly more than that of other employees.

Interquartile Range (IQR)

It is the difference between quartile of the third and first quartiles from the last. It can robust to the outliers.

Histograms

Histograms show the underlying frequency of a collection of discrete or continuous data on an interval scale.

Inferential Statistics

Inferential statistics allows you to draw conclusions about the population based on a sample set of data.

Hypothesis Testing

Hypothesis testing is a type of statistical reasoning that entails posing a question, gathering data, and then analysing what the data says about how to proceed. The null hypothesis is denoted by the symbol Ho and is the hypothesis that will be tested. We compare the null hypothesis to an alternative hypothesis, denoted by the letter Ha.

T Tests

When we just have a sample, rather than population statistics.

To calculate the population standard deviation, use the sample standard deviation. Because we only have samples, the T test is more prone to errors.

Z Score

The standard score, often known as the z score, is the distance between the observed value and the mean in units of standard deviations.

$$z = \frac{\bar{x} - \mu}{\sigma}$$

+Z – value is above mean.

-Z – value is below mean.

The distribution once converted to z- score is always same as that of shape of original distribution.

Chi Squared Test

To test the different categorical variables.

Correlation

Establish a connection between two variables. The letter r stands for it. The value might be anywhere between -1 and +1. As a result, 0 denotes the absence of a relationship.

Types

1. Supervised Learning

Supervised learning is a sort of algorithm that makes predictions based on a known dataset (called the training dataset). Input data and response values are included in the training dataset.

- Regression—which has a wide range of probable values. Eg- Exam marks
- Classification—which only has two options. Eg-Machine has a prediction of 0 or 1.

2. Unsupervised Learning

Unsupervised learning is the process of teaching a machine to learn from data that is neither classified nor labeled. The machine's objective here is to sort unsorted data into groups based on similarities, patterns, and differences without any prior data training.

- Clustering: A clustering problem is one in which you wish to uncover the data's intrinsic groupings, such as classifying clients based on their purchasing habits.
- Association: When solving an association rule learning problem, you want to find rules that describe huge chunks of your data, such as persons who buy X also buy Y.

Stages of Predictive Modelling

1. Problem definition
2. Hypothesis Generation
3. Data Extraction/Collection
4. Data Exploration and Transformation
5. Predictive Modelling
6. Model Development/Implementation

Problem Definition

Determine the most appropriate problem statement and, if possible, formulate the problem quantitatively.

Hypothesis Generation

Make a list of all possible variables that could affect the problem's goal. Personal bias and preferences should not be present in these variables.
The quality of the model is related to the quality of the hypothesis.

Data Extraction/Collection

Collect data from many sources and integrate it for model creation and exploration. We may discover new hypotheses while analysing data.

Data Exploration and Transformation

Data extraction is the process of retrieving data from a variety of sources for further data processing or storage.

Steps of Data Extraction

- Reading the Data. Eg. .ipynb files or .csv files
- Variable Identification
- Univariate Analysis
- Bivariate Analysis
- Missing Value Treatment
- Outlier Treatment
- Variable Transformation

Variable Treatment

It is the procedure for determining

1. whether a variable is Independent or Dependent.
2. Continuous or Categorical Variables.

Why do we perform variable identification?

1. Techniques like supervised learning require identification of dependent variable.
2. Different data processing techniques for categorical and continuous data.

Categorical variable- These are defined and stored like object.

Continuous Variable- These are defined and stored like integer or float.

Univariate Analysis

1. Concentrate on one variable at a time.
2. Make summary of the variable.
3. Make a sense of the summary to find insights, anomalies and so on.

Bivariate Analysis

1. When the empirical link between two variables is investigated.
2. When you want to see if the two variables are related to one another.
3. It aids in the detection and prediction of anomalies.

Missing Value Treatment

Causes of missing data

1. Non-response — For example, while collecting statistics on people's income, many people choose not to respond.

2. Data collection error. Faculty data, for example.

3. Data reading error.

Types :-

1. MCAR (Missing Completely at Random):- Missing values are unrelated to the variable in which they occur and other variables in the dataset.
2. MAR (Missing at Random):- Missing values have no relation to the variables in which they occur and variables other than those in which they occur.
3. MNAR (Missing not at random): Missing values have a relationship with the variable in

which they occur.

Missing values can be dealt with in a variety of ways.

1. Imputation

Continuous imputation using the mean, median, or regression mode.

The classification model is categorical-with mode.

2. Deletion

Delete rows or columns at a time. However, it results in loss of data and time.

Understanding the Execution Trading Benchmark

1. **TWAP (Time Average Weighted Average Price):** - This is the average of all transactions observed during a period of time. Regardless of their size, all trades are equally weighted. TWAP can be impacted significantly by even tiny operations at exorbitant prices. It is feasible to break large orders into tiny orders and disperse them equally over time to reduce the influence on the market or to prevent having a negative impact on the values of assets.

2. **VWAP (volume-weighted average price):** - The VWAP is the most extensively used and open standard. Aside from displaying how prices have changed over time, commercial execution is also displayed. The volume-weighted average price (VWAP) is the average of all volume-weighted exchanges. Small commercial sizes at extreme costs, unlike TWAP, have less of an impact on the benchmark and represent larger businesses.

3. **POV strategy (percentage of volume):** - Market participants specify the proportion of the volume (of the overall volume of the bag for this period) in which they want to participate in the POV strategy. The strategy's goal is to break down a large order into multiple smaller ones. Because market participants select the goal percentage, the size of a small order is proportional to the market's volume. The time duration for combining the volume and percentage of targets is specified by the participants. A large number of targets will work faster, but their impact on the market will be bigger.

4. **Spreading the calendar**: - One leg of the approach buys and sells a specific instrument in one flow, while the other leg buys and sells the same instrument in another flow. The goal of

trading is to profit from the various spreads that exist at various stages of the maturity process. As soon as the order is placed on the due date on which the trader is listed, the trader can trade on one leg (the two due days) and take a price on the other due date. Calendar spreads provide traders with a low-risk way to play the market because bidirectional transactions (BUY-SALE) are executed on the same basis, shielding the strategy from excessive market swings.

Outlier Treatment

The reasons for the outlier:-

1. Data entry Errors
2. Measurement Errors
3. Processing Errors
4. Change in underlying population

Outliers are classified into several categories.

1. Univariate

Only one variable is analysed in a univariate analysis to look for outliers. For example, consider a box plot of height and weight. We'll look for outliers in the weight.

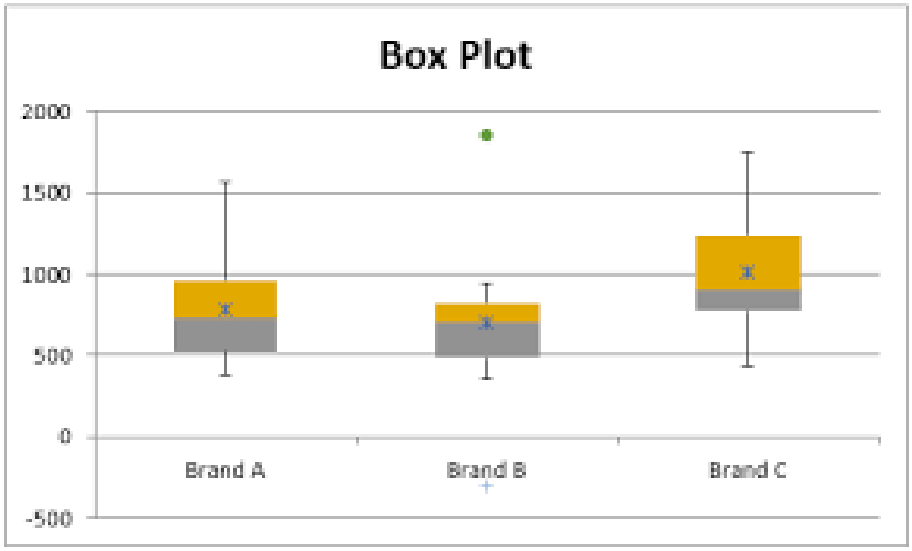
2. Bivariate

Looking for outliers in both variables.For example, consider a scatter plot graph of height and weight. Both will be analysed.

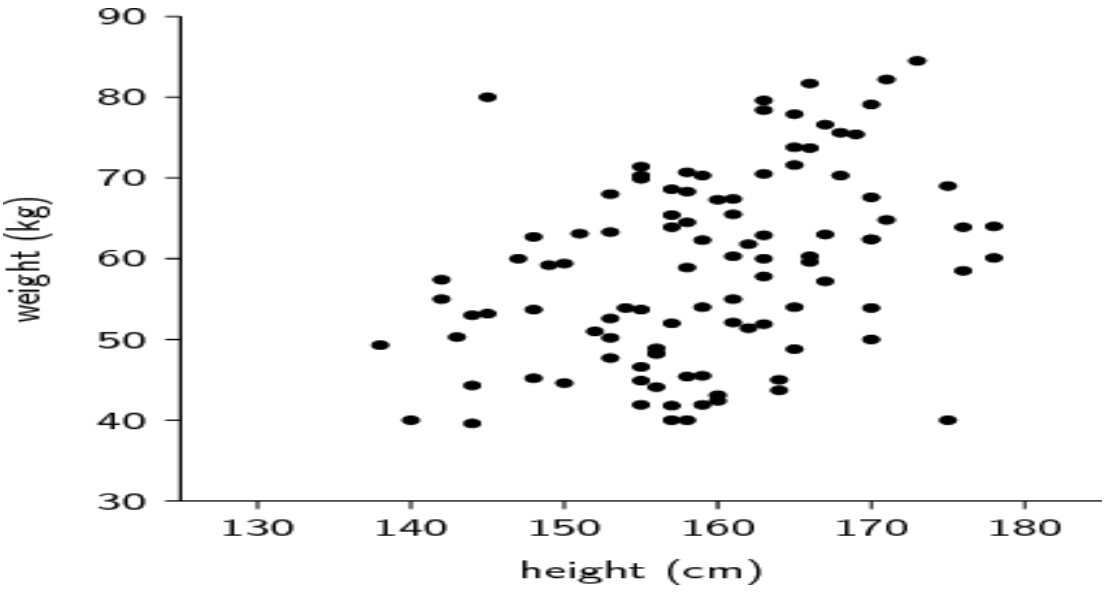
To Identify the Outlier

Graphical Method

a) Box Plot



b) Scattered Plot



Treatment of Outliers

1. Observations are deleted

2. Value transformation and binning

3. Adding missing values and other outliers

4. Handle them as individual Variable Transformations.

Is the procedure for

1. Replacing a variable with a function of that variable. Replacing a variable x with its log, for example.

2. We alter a variable's distribution or relationship with others. accustomed to –

Used to –

1. Change a variable's scale.
2. Converting non-linear interactions into linear ones.
3. Making a symmetric distribution from a skewed one.

Model Building

It is the process of developing a mathematical model for estimating and forecasting the future using historical data.

Eg- A retail wants to know the default behavior of its credit card customers. They want to know how likely each customer is to default in the following three months.

• The probability of default would range from 0 to 1.

• Assume a 10% default rate for every customer. In the next three months, the probability of each customer defaulting is 0.1.

It shifts the probability towards one of the extremes based on historical data qualities. A customer with violate income is more likely to default (or is on the verge of doing so).

A consumer with a good credit history over the last few years has a low risk of default (closer to 0).

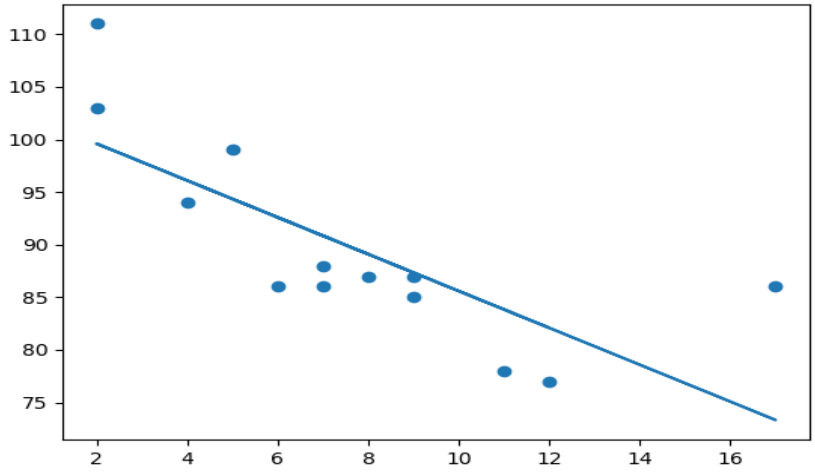
Steps in Model Building

1. Algorithm Selection
2. Training Model
3. Prediction / Scoring

Linear Regression

Linear regression is a statistical method for modelling a dependent variable's relationship with a set of independent variables.

The two variables are believed to be linearly connected. As a result, we look for a linear function. As a function of the feature or independent variable, this predicts the response value(y) as accurately as possible (x).

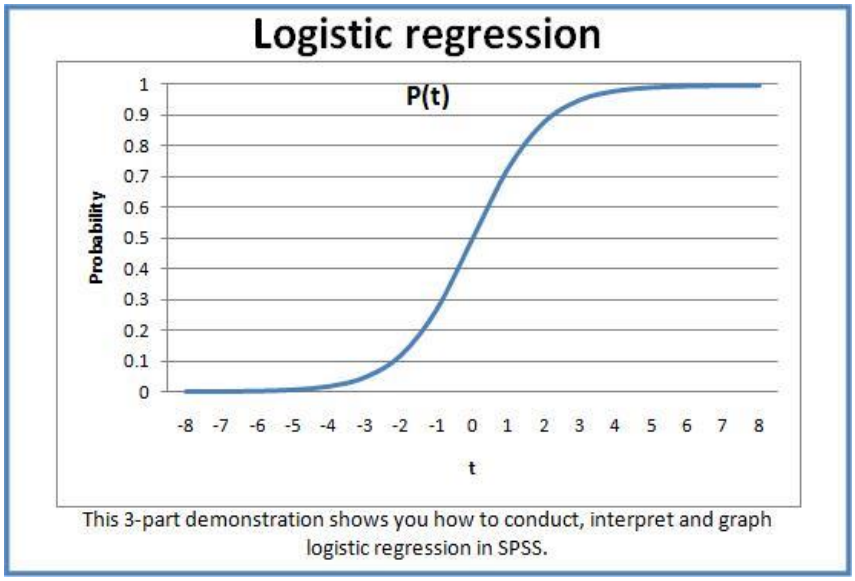


The equation of linear regression line is represented as:

$Y = a + bx$

Logistic Regression

Logistic regression is a statistical model that uses a logistic function to represent a binary dependent variable in its most basic form, though there are many more advanced variants.

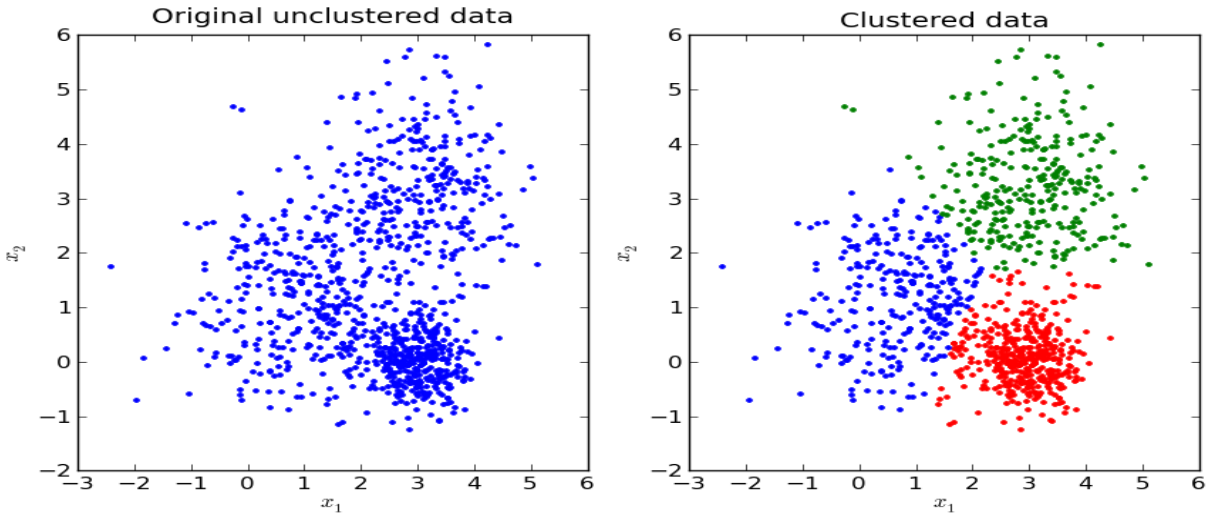


The equation of logistic regression is defined as:-

$C = -y (\log(y) – (1-y) \log(1-y))$

K-Means Clustering

When you have unlabeled data, K-means clustering is a sort of unsupervised learning that you can employ (i.e., data without defined categories or groups). The purpose of this technique is to locate groups in the data, with K representing the number of groups. Based on the attributes provided, the algorithm assigns each data point to one of K groups iteratively. Data points are grouped together based on how comparable their features are being similarity.



Predictive Analytics

Using historical data and qualities, we can forecast

Predicting stock price movement

1. Analyzing the past stock prices.
2. Analyzing the similar stocks.
3. Future stock price required.

When considering an investment, we must keep two things in mind: the upside and the downside. To put it another way, we should think about how much money we'll make if everything goes well and how much money we'll lose if the investments fail.

Types of Assets :-

Government bonds :- Government bonds have a 4-6 -percent average return, and there have been relatively few incidents of governments falling bankrupt and not repaying investors. As a result, there is some risk associated with this investment; it isn't risk-free, but the risk is well-managed.

Equity shares :- The rate of return on equity shares is higher. However, because diverse elements influence a company's share price, they are connected with considerably more frequent swings and price shifts.

The goal of finance isn't to maximise an investor's annual returns. It's all about making informed decisions that consider both risk and return, and optimising an investment portfolio's risk-return combination.

Calculating rates of return

1) **Simple Return** :- (Ending Price – Beginning Price) / (Beginning Price) When dealing with several assets in a short period of time, this is the best option.

2) **Log Return** :- Log (Ending Price / Beginning Price) When calculating the value of a single asset over time, this method is preferred.

Most investors have a number of equities in their portfolio, which is referred to as their **investment portfolio.**

Calculating a portfolio's rate of return is a simple and basic job. We have the rates of return of individual securities, and all we have to do now is multiply the rate of return of each security by its weight in the whole portfolio.

Rp = W1R1 + W2R2 …… +WnRn

A **market index** gives you a sense of how a certain stock market is doing. It is an excellent proxy for the market's overall development since it provides a large enough sample of the total number of stocks in the market.

A **stock index** represents the type of return you may expect in a given market if you invest in a well-diversified portfolio.

In the field of finance, variability is quite essential. It is the most accurate risk indicator we have. A stock that is volatile is considerably more likely to diverge from its historical performance and negatively surprise investors.

$$s^2 = \frac{\sum (X - \overline{X})^2}{N - 1}$$

Sample Variance

When attempting to evaluate risk linked with dispersion in the anticipated outcome, commonly used statistical measures such as variance and standard deviation can be extremely useful. The variance and standard deviation of a security are used to measure this dispersion.

It's logical to assume that **common factors influence the values of shares on a stock exchange**. The development of the economy is the most obvious example. In general, positive macroeconomic conditions make it easier for all businesses to do business. The status of the economy has an impact on the value of a company's stock. Different industries, on the other hand, are influenced in different ways. In times of crisis, certain industries fare

better than others.

1. **Perfect Correaltion:-** 1 is equal to the first variable explains the total variability of the second variable.
2. **Negative Correlation**:- It could be a perfect negative correlation of -1 or a considerably more likely imperfect negative correlation with a value between -1 and 0.
3. **Neural Correlation**:- A correlation of 0 between two variables indicates that they are unrelated.

Calculating Correlation

$$Correlation = \frac{Cov\ (x,y)}{\sigma x * \sigma y}$$

Calculating Covariance

$$COV(X,Y) = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{n-1}$$

If a portfolio consists of two stocks, the risk will be determined by the variances of the two stocks as well as their correlation.

$$(w_1\sigma_1 + w_2\sigma_2)^2 = w_1^2\sigma_1^2 + 2w_1\sigma_1 w_2\sigma_2\rho_{12} + w_2^2\sigma_2^2$$

Systematic Risk:- This is the unpredictability that characteristics the entire market. The day-to-day fluctuations in stock prices are known as systematic risk, and it is produced by events that influence all companies.

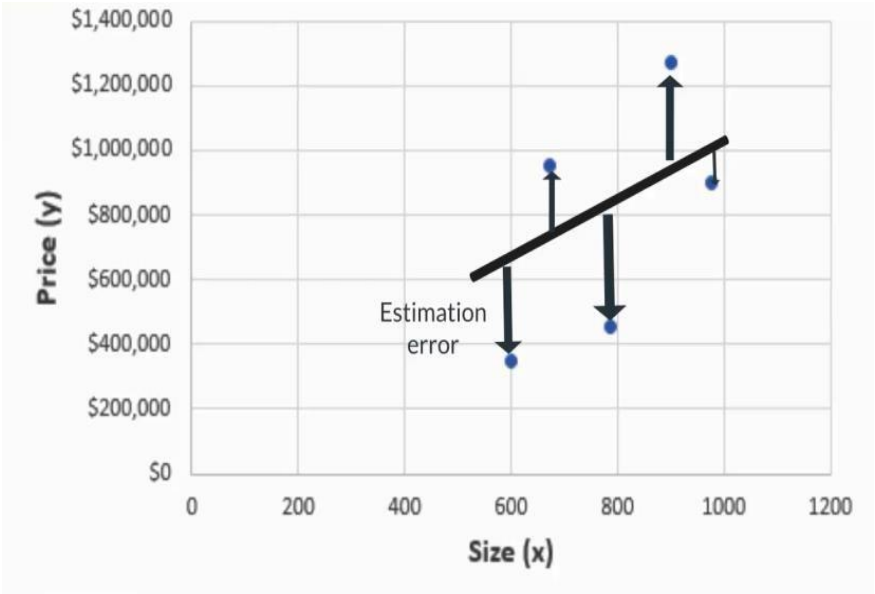
Unsystematic Risk:- These are company-specific, even industry-specific risks that diversification might help to lower risk.

Regression Analysis

Regression Analysis:- In the field of finance, regression analysis is one of the most commonly utilized tools. It measures the relationship between one or more explanatory variables, often known as independent variables, and a dependent variable.

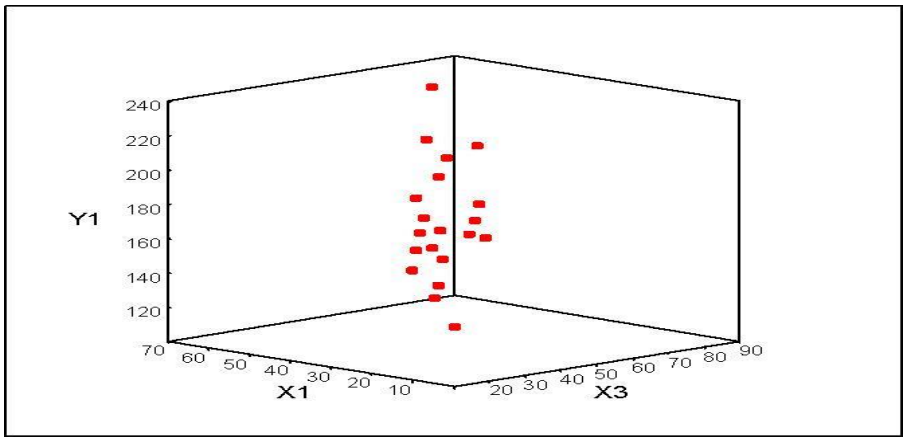
Types of Regression Analysis

1. **Simple Regression: -** The existence of a linear relationship between the two variables is assumed in regression analysis. The best fit is a single straight line, which can help us characterise the relationship between all of the data points in the plot. It can be calculated by Yi=β0+β1 $X1$+εi.



Minimizing the estimation error (Simple Regression)

2. **Multivariate Regression:-** We can improve the explanatory power of the regression equation by including more variables in it. This will give us a better view of the whole picture of events that influence the development of the variable we're trying to forecast.



Minimizing the estimation error (Multivariate Regression)

Statisticians have developed an easy-to-understand tool. It's known as R2. To comprehend R2, we must consider the overall variability of the data. TSS gives you a sense of the data's variability. The calculation for x variance is the same as before, except N – 1 is not included.

The total variability of the data can also be broken down into the following components:

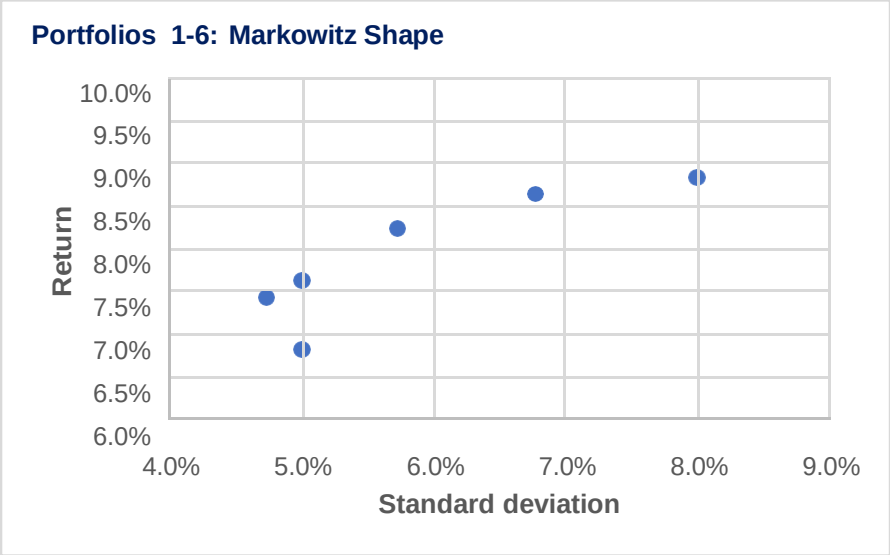
TSS = SSE + SSR

where SSE is for the sum of squares, and SSR stands for the sum of squared residuals.

Markowitz Model

Markowitz demonstrated the existence of a set of portfolios that maximizes an investor's return for the amount of risk they are willing to take.

One of the most important takeaways from his research is that multi-security investments should not be assessed independently, but rather as part of a portfolio, and that financiers must understand how different securities in a portfolio interact with one another.



Investors can optimise their returns without taking on more risk, according to Markowitz, by combining low-correlation assets.

The Capital Asset Pricing Model of William Sharpe includes the following components:

- **Market portfolio :-** This portfolio contains all potential investments in the world (including bonds and equities), and its risk-return combination outperforms that of any other portfolio.
- **Risk Free Asset :-** An absolutely risk-free investment (zero standard deviation). It has a good rate of return, but there is no risk involved.
- **Beta Coefficient :-** It aids us in quantifying the link between a securities and the total market portfolio.
- **Capital Market Line :-** The Capital Market Line is a tangent to the efficient frontier that connects the risk-free rate with the efficient frontier. The market portfolio is the point where the Capital Market Line intersects the efficient frontier.

Capital Asseet Pricing Model formula:-

The expected return of a security is equal to the risk-free asset's return + beta times the market's expected return minus the risk-free asset's return.

Risk-free: The smallest amount of profit an investor may expect from a given investment.

The beta coefficient indicates how risky a certain asset is in comparison to the market.

When purchasing the Market portfolio, an investor should expect to get a market risk premium.

Sharpe Ratio Formula

Rational investors aim to maximise their rate of return while minimising their investment's risk, therefore they require a risk-adjusted return measure, a tool that allows them to analyse different securities and invest in the ones that would produce the highest return for a given degree of risk.

This is how William Sharpe got his renowned Sharpe ratio from. It's a great technique to compare stocks and portfolios and determine which is better in terms of risk and return.

$$S = \left(\frac{R_p - R_f}{\sigma_p} \right)$$

Types of Investment strategies:-

1. **Passive Investing:-** It consist of purchasing an asset portfolio and holding it for the long term, independent of short-term macroeconomic events.
2. **Active Investing :-** Trading on a regular basis based on macroeconomic and company-

specific trends.

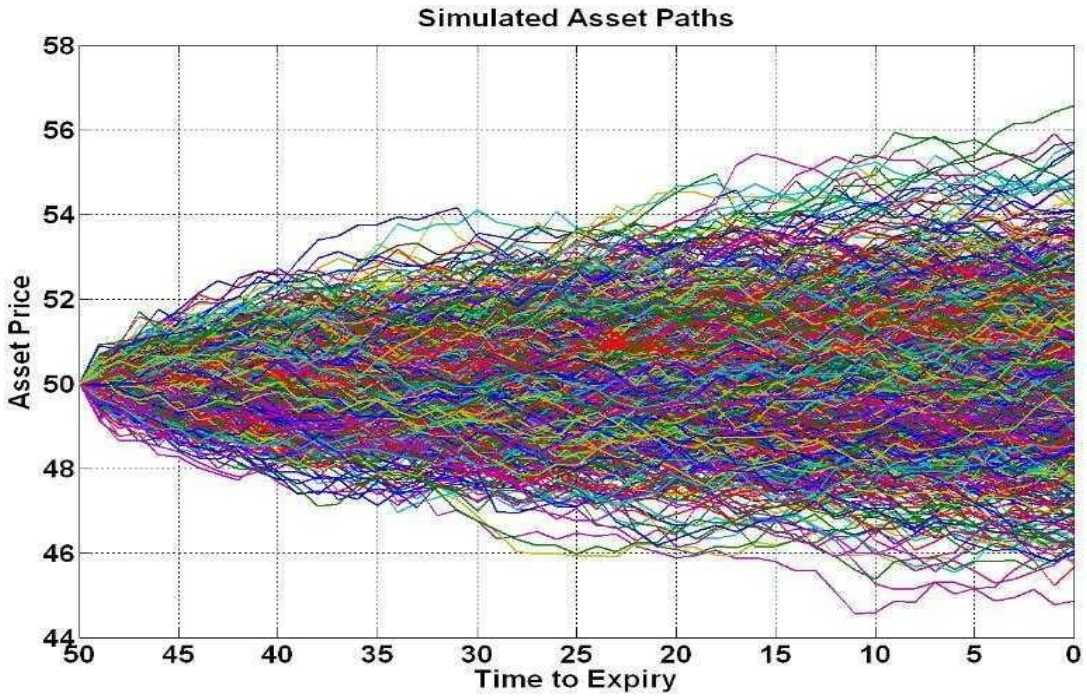
3. **Arbitage Investing:-** Find pricing differences in the market and take advantage of them to earn without taking any risks.
4. **Value Investing:**- Invest in certain companies in the hopes of outperforming their competitors.

Monte Carlo Simulations

We want to see all of the numerous possible outcomes of a future event when we perform a Monte Carlo simulation. What occurs in real life is only one of the conceivable outcomes of any given situation.

A Monte Carlo simulation comes in handy in this situation. We can develop a simulation — a new set of fictional but sensible data – by using previously collected data. These realizations are created by looking at the historical data distribution and estimating the mean and variance.

Such data is useful because it allows us to consider a good proxy for the probability of various events and can assist us in making an informed decision.



Derivatives were first used as a hedging tool. The majority of companies interested in purchasing these contracts were concerned about protecting their investment. Financial institutions, on the other hand, brought a tremendous deal of innovation to the scene over time, employing so-called financial engineering and introducing new types of derivatives.

- **Forward :-** A forward contract is utilised when two parties agree that one will sell an underlying asset to the other at a later date. The asset's price is agreed upon

ahead of time.

- **Future:-** Futures are highly standardised forward contracts that are usually agreed upon in a market. The difference between futures and forwards is the amount of standardisation and the involvement of a clearing house - the transaction takes place on the open market, and the counterparties are strangers.
- **Swaps:-** Swap contracts are financial derivatives in which two parties agree to exchange cash flows dependent on the performance of an underlying asset at a future date. An interest rate, a stock price, a bond price, a commodity price, and so on can all be used as the underlying asset.
- **Options:-** A strike price is the price at which an option contract's owner can purchase or sell an underlying asset. The owner of the option contract has the choice to buy or sell the asset at the specified price, but he can also choose not to do so if the asset's price is unfavorable.

Prescriptive Analysis

Prescriptive analytics is based on artificial intelligence techniques such as machine learning, which is the ability of a computer programme to understand and advance from data without extra human input while adapting. Machine learning allows for the processing of today's massive amounts of data. As new or extra data becomes available, computer programmes automatically change to make advantage of it, in a far faster and more complete manner than human capacities could handle.

Both descriptive and predictive analytics are related to prescriptive analytics. Prescriptive analytics strives to discover the optimal solution or outcome among multiple options, given the available parameters, whereas descriptive analytics aims to provide insight into what has happened and predictive analytics helps model and anticipate what might happen.

Prescriptive analytics can also recommend decision options for how to capitalize on a future opportunity or reduce a future risk, as well as highlight the consequences of each option. In practice, prescriptive analytics can analyses fresh data continuously and automatically to enhance forecast accuracy and provide improved choice possibilities.

The prescriptive analysis includes various points like:-

- Prescriptive analytics employs machine learning to assist businesses in making decisions based on the predictions of a computer programed.

- Prescriptive analytics is use to help in predictive analytics, which uses data to forecast short-term outcomes.

- Prescriptive analytics, when utilised correctly, can assist organisations make decisions based on facts and probability-weighted estimates, rather than rash decisions based on instinct.

Part C

Learning Outcome

After completing the training, I have learnt little fundamental and technical process through Data Analytics and how it works in the daily life.

1. **Market Risk :-** To analyses, the risk arising from price fluctuations in the securities in which the company is interested), sensitivity analysis, VaR analysis, and stress test can be used.

2. **Credit Risk:-** Risks originating from counterparty obligations and payment default) A dashboard technique can be used to assess these risks. The ratings of independent rating organizations can be utilized to define the thresholds for each counter-party in this technique.

3. **Financing Risk:-** It is possible to evaluate the risk associated with the need to fund the monetary needs of typical commercial transactions) as well as the risk of yield curve adjustments.

4. **Regulatory Risk:-** Risk posed by laws and regulatory developments) as well as risk posed by pooling best efforts to predict future regulatory developments based on present events and context. It can be assessed.

5. **Liquidity Risk:-** Risk of loss as a result of being unable to trade due to a lack of liquidity to trade at favorable levels. There is also a loss owing to the requirement to pay an ask-spread during an exchange: when exchanging significant amounts in one direction, the market price will reverse in the opposite direction of the exchange.

6. **Operational Risk:-** Mechanical, criminal, natural, terrorist, and other systems (Risks related to normal daily operations, adequate functioning of technological systems, physical security, political risks, etc.). You can also protect yourself against threats like fraud. It can be handled by enhancing the system's general resilience (such as legal risk), boosting the system's capacity to manage demand, and analyzing the maximum order flow before an error is detected.

The Importance of Buying and Selling stocks in volume

- The required delivery capacity for volume trading must be carefully considered.
- Traders looking for increased buy or sell orders are known as volume traders. They

also take present and potential price movements into consideration.

- Increasing trading volume improves orders in general. Traders are opening new positions as a result of the strong volume trends.

- If your cash flow and trading volume both fall, you'll notice a "stock divergence," which means it's probably time to sell.

Project Delivearables

- Develop relevant programming abilities on Python.
- Demonstrate knowledge of statistical data analysis.
- Develop the ability to create and evaluate data-driven models.
- Use professional statistical software to do statistical analysis.
- Demonstrate data management expertise.
- Use data analysis principles and methodologies to solve problems in real-world situations and effectively convey the results.

Bibliography

https://www.mastersindatascience.org/learning/what-is-data-analytics/

https://corporatefinanceinstitute.com/resources/knowledge/finance/regression-analysis/

https://quantpedia.com/markowitz-model/

https://en.wikipedia.org/wiki/Monte_Carlo_method

https://in.tradingview.com/

https://www.anaconda.com/

Gateway Group

Date: 8th July, 2021

To Whomsoever It May Concern

This to certify that **Mr. Anshuman Singh Deora** from **Institute of Management, Nirma University** has successfully completed his summer internship with us from **05th May, 2021 to 05th July, 2021.**

He was exposed to **Data Analytics with Python** on the project 'Finance and Investment Strategies in Indian Market'. The project involved fundamental and technical analysis such as Filtering Data, Data Gathering, Data Extraction, Data Visualization, Regression Analysis and calculating and comparing rates of return on python.

We would sincerely like to acknowledge the amazing work he has accomplished during this project, working so diligently setting high standards of genuineness and reliability.

During the period of internship, we found him very much passionate, interested to learn, willing to put his best efforts and get into the depth of the subject to understand it better for implementation and execution.

We wish him all the best in his future endeavours.

Note: - As per the company guidelines, confidential information relating to code cannot be shared with external parties.

With Best Regards,

For, Gateway TechnoLabs Pvt. Ltd.
(A part of Gateway Group of Companies)

Pratik Mehta

Authorized Signatory
Pratik Mehta

Gateway TechnoLabs Private Limited.
B - 81 Corporate House, S.G. Highway, Bodakdev, Ahmedabad - 380054 Gujarat.
Fax: +91-79-26858591 | | **Tel:** +91-79-26852554/5/6
Email: gateway@thegatewaycorp.com | www.thegatewaycorp.com | **CIN:** U72200GJ2001PTC39707