

DATA ANALYSIS & VISUALIZATION

A Project-II Report

Submitted in partial fulfillment of requirement of the

Degree of

**BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE &
ENGINEERING**

BY

Aayush Khandelwal

EN17CS301003

Under the Guidance of



Department of Computer Science & Engineering

Faculty of Engineering

MEDI-CAPS UNIVERSITY, INDORE- 453331

April 2021

Report Approval

The project work “**Data Analysis & Visualization**” is hereby approved as a creditable study of an engineering/computer application subject carried out and presented in a manner satisfactory to warrant its acceptance as prerequisite for the Degree for which it has been submitted.

It is to be understood that by this approval the undersigned do not endorse or approved any statement made, opinion expressed, or conclusion drawn there in; but approve the “Project Report” only for the purpose for which it has been submitted.

Internal Examiner

Ms Anusha Jain

Computer Science & Engineering

Medi-Caps University, Indore

External Examiner

Name:

Designation

Affiliation

Declaration

We hereby declare that the project entitled “**Data Analysis & Visualization**” submitted in partial fulfillment for the award of the degree of Bachelor of Technology of Computer Applications in ‘Computer Science & Engineering’ completed under the supervision of **Ms Anusha Jain, Assistant professor, Computer Science & Engineering**, Faculty of Engineering, Medi-Caps University Indore is an authentic work.

Further, we declare that the content of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted to any other Institute or University for the award of any degree or diploma.

**Aayush
Khandelwal**

(EN17CS301003)

Certificate

I, **Ms. Anusha Jain** certify that the project entitled “**Data Analysis & Visualization**” submitted in partial fulfillment for the award of the degree of Bachelor of Technology of Computer Science and Engineering by **Aayush Khandelwal** is the record carried out by them under my guidance and that the work has not formed the basis of award of any other degree elsewhere.

Ms.Anusha Jain

Computer Science & Engineering

Medi-Caps University, Indore

Mr.Jayesh Soni

Bharat Financial Inclusion Ltd.

Dr. Suresh Jain

Head of the Department

Computer Science & Engineering

Medi-Caps University, Indore

Acknowledgement

I would like to express my deepest gratitude to Honorable Chancellor, **Shri R C Mittal**, who has provided me with every facility to successfully carry out this project, and my profound indebtedness to **Prof. (Dr.) Sunil K Somani**, Vice Chancellor, Medi-Caps University, whose unfailing support and enthusiasm has always boosted up my morale. I also thank **Prof. (Dr.) D K Panda**, Dean, Faculty of Engineering, Medi-Caps University, for giving me a chance to work on this project. I would also like to thank my Head of the Department **Dr. Suresh Jain** for his continuous encouragement for betterment of the project.

We express our heartfelt gratitude to our **Project Guide, Ms Anusha Jain**, Assistant professor, Computer Science & Engineering, Medi-Caps University, Indore (M.P.) without whose continuous help and support, this project would never have reached to the completion.

We would also like to thank our **Mentor Mr Kailash Bandhu and Mr. Abhishek Sable**, Assistant professor, Computer Science & Engineering, Medi-Caps University, Indore (M.P.) for continuous assisting us and providing valuable guidance.

Aayush Khandelwal

B.Tech. IV Year

EN17CS301003

Department of Computer Science & Engineering

Faculty of Engineering

Medi-Caps University, Indore

ABSTRACT

This document is a review report on the the project made in the field of computer engineering to perform data analysis and visualization to help gather insights about a business organization to improve its business operations. The report proposed the results and solutions on the limited implementation of the various techniques that are introduced in the project.

Furthermore, the paper states the overview of the observations made by the authors in order to help further optimization in the mentioned field to achieve the utility at a better efficiency .

Keywords: Data analysis , Visualization , Matplotlib, Power BI , Report.

Table of Contents

		Page No.
	Report Approval	ii
	Declaration	iii
	Certificate	iv
	Acknowledgement	v
	Abstract	vi
	Table of Contents	7
	List of figures	8
	Abbreviations	9
Chapter 1	Introduction	10
1.1	Introduction	11
1.2	Objective	12
1.3	Scope	12
1.4	Problem Statement	13
Chapter 2	Software Requirement Analysis	14
2.1	System Feasibility	15
2.2	Platform Specification	15-16
Chapter 3	Loan approval prediction	17
3.1	Introduction	18
3.2	Methodology	18
3.3	Phases of Data Analysis	19-20
3.4	Steps Involved	21-24
3.5	Insights	25
Chapter 4	Power BI	26
4.1	Introduction	27
4.2	What is Power BI?	27-28
4.3	Building blocks of Power BI	29-31
Chapter 5	Limitation	32-33
Chapter 6	Future Scope	34-35
Chapter 7	Conclusion	36-37
Chapter 8	References	38

List of Figures

S. No.	Figure Name	Page No.
1	Figure 3.1 (Univariate analysis)	22
2	Figure 3.2 (Univariate analysis)	22
3	Figure 3.3 (Bivariate analysis)	23
4	Figure 3.4 (Bivariate analysis)	23
5	Figure 3.5 (Structure of confusion matrix)	24
6	Figure 4.1(Power BI Architecture)	28
7	Figure 4.2(Visualization)	29
8	Figure 4.3(Report)	30
9	Figure 4.4(Power BI Interface Look)	31
10	Figure 4.5(Dashboard view)	31

Abbreviations

Acronym	Full Form
EDA	Exploratory Data Analysis
BI	Business Intelligence

CHAPTER 1

INTRODUCTION

1.1 Introduction

Data is all around us and everything that we do results in new data. Every kind of electronic message, which we either receive or send like withdrawing money from a bank, every website that we visit contributes to the storage of data. In today's world data is a commodity whose value is incalculable.

Data analytics is the key to driving productivity, efficiency and revenue growth. The results from analysing data sets is going to tell an organisation where they can optimise, which processes can be optimised or automated. In this way cost effectiveness is increased as areas that are hording a company's finances unnecessarily can be identified and decisions can be made around technologies that can be put in place to reduce operational and production costs. Data Analytics has a key role in improving business as it is used to gather hidden insights, generate reports, perform market analysis, and improve business requirements. many tools have emerged with various functionalities for this purpose.

Data analytics is a broad field. There are four primary types of data analytics: descriptive, diagnostic, predictive and prescriptive analytics. Each type has a different goal and a different place in the data analysis process. These are also the primary data analytics applications in business.

- **Descriptive analytics**: It helps answer questions about what happened. These techniques summarize large datasets to describe outcomes to stakeholders. By developing key performance indicators these strategies can help track successes or failures. This process requires the collection of relevant data, processing of the data, data analysis and data visualization. This process provides essential insight into past performance.
- **Diagnostic analytics**: It helps answer questions about why things happened. These techniques supplement more basic descriptive analytics. They take the findings from descriptive analytics and dig deeper to find the cause. The performance indicators are further investigated to discover why they got better or worse.
- **Predictive analytics**: It helps answer questions about what will happen in the future. These techniques use historical data to identify trends and determine if they are likely to reoccur. Predictive analytical tools provide valuable insight into what may happen in the future and its techniques include a variety of statistical and machine learning techniques, such as: neural networks, decision trees, and regression.
- **Prescriptive analytics**: It helps answer questions about what should be done. By using insights from predictive analytics, data-driven decisions can be made. This allows businesses to make informed decisions in the face of uncertainty. Prescriptive analytics techniques rely on machine learning strategies that can find patterns in large datasets. By analyzing past decisions and events, the likelihood of different outcomes can be estimated. These types of data analytics provide the insight that businesses need to make effective and efficient decisions. Used in combination they provide a well-rounded understanding of a company's needs and opportunities.

1.2 Objective

- Gather Hidden Insights – Hidden insights from data are gathered and then analyzed with respect to business requirements.
- Better customer targeting and ensuring growth- By understanding clients more fully, and by using analytics of their transactions and trading activities, company can be sure that they are delivering the best services for what their customers need, resulting in higher levels of retention and acquisition.
- Generate Reports – Reports are generated from the data and are passed on to the respective teams and individuals to deal with further actions for a high rise in business.
- Enhancing risk assessment - As banks will be able to assess the risk profiles of their credit applicants in much greater detail, they will also be able to improve their credit assessments.
- Perform Market Analysis – Market Analysis can be performed to understand the strengths and weaknesses of competitors.

1.3 Scope

- Data analysis is the process of developing answers to questions through the examination and interpretation of data. The basic steps in the analytic process consist of identifying issues, determining the availability of suitable data, deciding on which methods are appropriate for answering the questions of interest, applying the methods and evaluating, summarizing and communicating the results.
- Analytical results underscore the usefulness of data sources by shedding light on relevant issues. Data analysis also plays a key role in data quality assessment by pointing to data quality problems in a given survey. Analysis can thus influence future improvements to the survey process.
- Data analysis is essential for understanding results from surveys, administrative sources and studies, for providing information on data gaps, for designing and redesigning surveys, for planning new statistical activities and for formulating quality objectives.

1.4 Problem Statement

Organizing Data- Basically, huge amount of raw data is been generated every day. These data includes transaction made per day, location of customers across the country, date of transaction, amount etc. When this much amount of data is just stored in data base, no worthwhile information can be drawn, but when advocated by data analytics these raw data can be converted into some useful information which can help in providing insight to improve customer service, and build product faster. With the process of data mining, it provides the ability to react even before any financial failures and finding ways to monetize it. The idea is to make sense of the data you have, to analyze it and share better business prospects in the near future and how you're going to do it, is with the concepts of analytics. Data Science involves extraction of trends, patterns and useful information from a set of existing data which will be of no use if not analyzed. It is a kind of business intelligence that is now used for gaining profits and making better use of resources. This can also help in improving managerial operations and leverage organizations to next level.

CHAPTER 2

SYSTEM REQUIREMENT

ANALYSIS

2.1 System Feasibility

2.1.1 Technical Feasibility:

- The project covers all the scope of practicality and is technically feasible.
- It will be built considering industry standards, and using the latest technologies which will ensure its smooth functioning.
- The project will run on users devices over internet.

2.1.2 Economical Feasibility:

- The hardware requirements to access this project are appropriate computing device.

2.2 Platform Specification

2.2.1 Hardware Requirements:

- **Laptop:**Used to run our code and perform data analysis and visualization using certain tools.
- 4 GB or more RAM

2.2.2 Software Requirements:

- **Python** :Python is an object-oriented programming language created by Guido Rossum in 1989. Python is open source, interpreted, high level language and provides great approach for object-oriented programming. It is one of the best language used by data scientist for various data science projects/application. Python provide great functionality to deal with mathematics, statistics and scientific function. It provides great libraries to deals with data science application.
- **Power BI** is a Data Visualization and Business Intelligence tool that converts data from different data sources to interactive dashboards and BI reports. Power BI suite provides multiple software, connector, and services - Power BI desktop, Power BI service based on Saas, and mobile Power BI apps available for different platforms. These set of services are used by business users to consume data and build BI reports.

2.2.3 Tools and Libraries used:

- **Jupyter notebook:** The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.
- **Matplotlib** is another useful Python library for Data Visualization. Descriptive analysis and visualizing data is very important for any organization. Matplotlib provides various methods to visualize data in a more effective way. Matplotlib allows to quickly make line graphs, pie charts, histograms, and other professional grade figures. Using Matplotlib, one can customize every aspect of a figure. Matplotlib has interactive features like zooming and panning and saving the Graph in graphics format.
- **Sklearn** is a Python library for machine learning. Sklearn provides various algorithms and functions that are used in machine learning. Sklearn is built on NumPy, SciPy, and matplotlib. Sklearn provides easy and simple tools for data mining and data analysis. It provides a set of common machine learning algorithms to users through a consistent interface. Scikit-Learn helps to quickly implement popular algorithms on datasets and solve real-world problems.
- **Pandas** is one of the most popular Python libraries for data manipulation and analysis. Pandas provides useful functions to manipulate large amounts of structured data. Pandas provides the easiest method to perform analysis. It provides large data structures and manipulating numerical tables and time series data. Pandas is a perfect tool for data wrangling. Pandas is designed for quick and easy data manipulation, aggregation, and visualization.
- **Numpy** is a Python library that provides mathematical functions to handle large dimension arrays. It provides various methods/functions for Array, Metrics, and linear algebra.
NumPy stands for Numerical Python. It provides lots of useful features for operations on n-arrays and matrices in Python. The library provides vectorization of mathematical operations on the NumPy array type, which enhances performance and speeds up the execution. It's very easy to work with large multidimensional arrays and matrices using NumPy.

CHAPTER 3

LOAN APPROVAL

PREDICTION

3.1 Introduction

Distribution of the loans is the core business part of almost every banks. The main portion the bank's assets is directly came from the profit earned from the loans distributed by the banks. The prime objective in banking environment is to invest their assets in safe hands where it is. Today many banks/financial companies approves loan after a regress process of verification and validation but still there is no surety whether the chosen applicant is the deserving right applicant out of all applicants. Through this system we can predict whether that particular applicant is safe or not and the whole process of validation of features is automated by machine learning technique.

Loan Prediction is very helpful for employee of banks as well as for the applicant also. It can provide special advantages to the bank. The Loan Prediction System can can automatically calculate the weight of each features taking part in loan processing and on new test data same features are processed with respect to their associated weight .A time limit can be set for the applicant to check whether his/her loan can be sanctioned or not. Loan Prediction System allows jumping to specific application so that it can be check on priority basis.

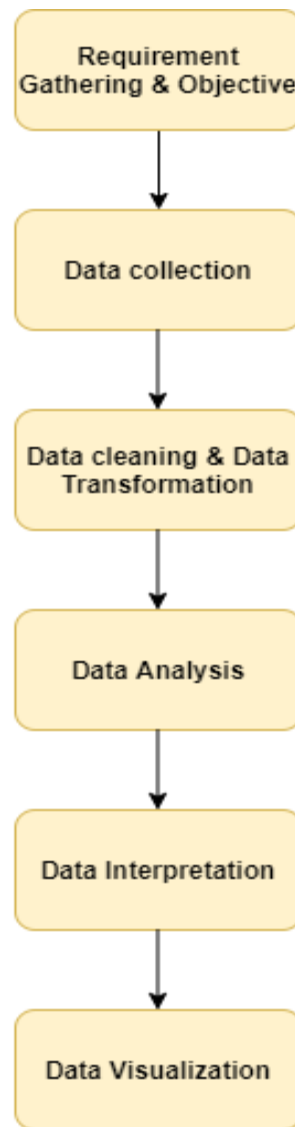
3.2 Methodology

The proposed model focuses on predicting the credibility of customers for loan repayment by analyzing their behavior. The input to the model is the customer behavior collected. On the output from the classifier, decision on whether to approve or reject the customer request can be made. Using different data analytics tools loan prediction and there severity can be forecasted. In this process it is required to train the data using different algorithms and then compare user data with trained data to predict the nature of loan. To extract patterns from a common loan approved dataset, and then build a model based on these extracted patterns. The training data set is now supplied to machine learning model; on the basis of this data set the model is trained. Every new applicant details filled at the time of application form acts as a test data set. After the operation of testing, model predict whether the new applicant is a fit case for approval of the loan or not based upon the inference it conclude on the basis of the training data sets. To extract important information and predict if a customer would be able to repay his loan or not.

3.3 Phases of Data Analysis:

Data Analysis consists of the following phases:

- **Data Requirement Gathering** - First of all, you have to think about why do you want to do this data analysis? All you need to find out the purpose or aim of doing the Analysis of data.
- **Data Collection** - Guided by the requirements identified, it's time to collect the data from sources. Sources include, surveys, interviews, questionnaires, direct observation, and focus groups, transactions, location of employees, date of transaction etc. And to make sure to organize the collected data for analysis.
- **Data Cleaning** - Now whatever data is collected may not be useful or irrelevant to your aim of Analysis, hence it should be cleaned. The data which is collected may contain duplicate records, white spaces or errors. The data should be cleaned and error free. This phase must be done before Analysis because based on data cleaning, your output of Analysis will be closer to your expected outcome.
- **Data Analysis** - Once the data is collected, cleaned, and processed, it is ready for Analysis. As you manipulate data, you may find you have the exact information you need, or you might need to collect more data. During this phase, you can use data analysis tools and software which will help you to understand, interpret, and derive conclusions based on the requirements.
- **Data Interpretation** - After analyzing your data, it's finally time to interpret your results. You can choose the way to express or communicate your data analysis either you can use simply in words or maybe a table or chart. Then use the results of your data analysis process to decide your best course of action.
- **Data Visualization**- Data visualization is very common in your day to day life; they often appear in the form of charts and graphs. In other words, data shown graphically so that it will be easier for the human brain to understand and process it. Data visualization often used to discover unknown facts and trends. By observing relationships and comparing datasets, you can find a way to find out meaningful information.



Flowchart (Phases of Data Analytics)

3.4 Steps Involved :

1. Firstly let us understand about the dataset. It includes 12 independent variables like 'Loan-ID', 'Gender', 'Married', 'Dependents', 'Education', 'Self-employed', 'Applicant Income', 'CoapplicantIncome', 'Loan Amount', 'Loan_Amount_Term', 'Credit History', 'Property Area' and 1 target variable i.e. Loan status.
2. Now for the project import all the important libraries like numpy , pandas, matplotlib, seaborn etc for further data analysis and visualization.
3. Converting the raw data set into csv format and then loading the csv file through pandas. `df = pd.read_csv("train.csv")`
4. Then format of the dataset can be understood by using functions like `head()`, `tail()`, `describe()`, `info()`.
5. Dataset contain different types of variables :
 - **Categorical Variables:** Categorical variables are those data fields that can be divided into definite groups. In this case, Gender(Male OR Female), Married(Yes Or No), Education(Graduate Or Not Graduate) etc.
 - **Ordinal variables** are the ones that can be divided into groups, but these groups have some kind of order. Like, high, medium, low. Dependents field can be considered ordinal since the data can be clearly divided into 4 categories : 0, 1, 2, 3+ and there is a definite ordering also. Same is the case with Property_Area (Urban, Semi-urban Or Rural).
 - **Numerical or Continuous Variables:** Numerical variables are those that can take up any value within a given range. For example, applicantIncome, CoapplicantIncome, Loan_Term, Loan_Amount.
6. Then we perform data pre-processing which is a technique to convert raw data set into a clean data set for better analysis. For this firstly we need to identify and correctly handle the missing values, failing to do this, you might draw inaccurate and faulty conclusions and inferences from the data.
 - We can do that using the `df.isnull().sum()` method which returns the column names along with the number of NaN values in that particular column.
 - The easiest way to solve this problem is by dropping the rows or columns that contain null values but it can lead to loss of features which might be important.
 - Therefore for numerical terms we fill the missing values with the mean of all the values present in that column.
 - For categorical terms we fill the missing values using the mode of that column.

7. Then we perform exploratory data analysis (EDA). EDA is a very crucial step. It gives us a glimpse of what our data set is all about, its uniqueness, its anomalies and finally it summarizes the main characteristics of the dataset for us.

- **Univariate analysis** involves analysis of one variable at a time. Let's say "Gender" then we will analyze only the "Gender" field in the dataset. The analysis is usually summarized in the form of count.

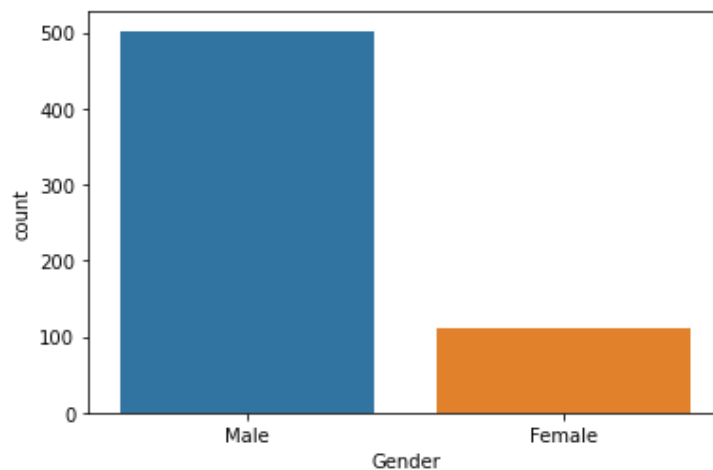


Figure 3.1

Similarly we can do it for variable 'Education'.

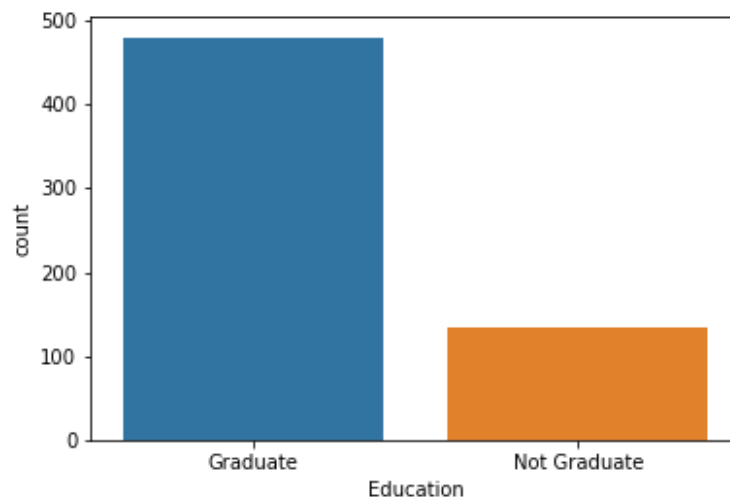


Figure 3.2

- Now let's find some relationship between two variables, particularly between the target variable "Loan_Status" and a predictor variable from the dataset. Formally, this is known as bivariate analysis. **Bivariate analysis** is finding some kind of empirical relationship between two variables.

For eg we draw Education wrt Loan status plot using crosstab Function to compare these 2 variables.

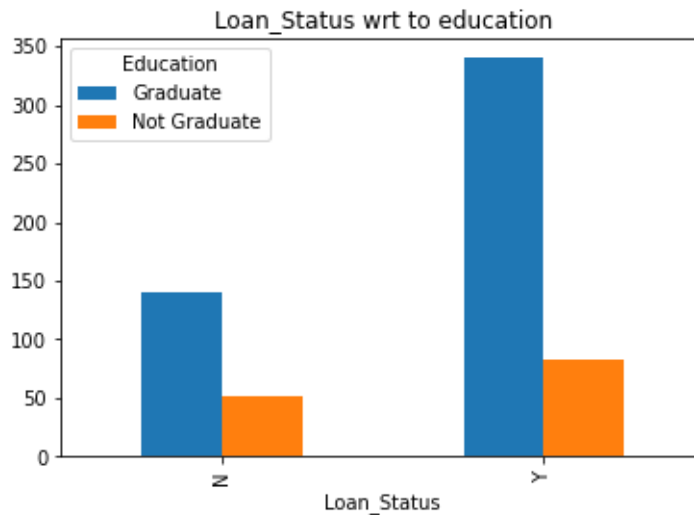


Figure 3.3

Similarly we can do comparison between Gender and Loan status.

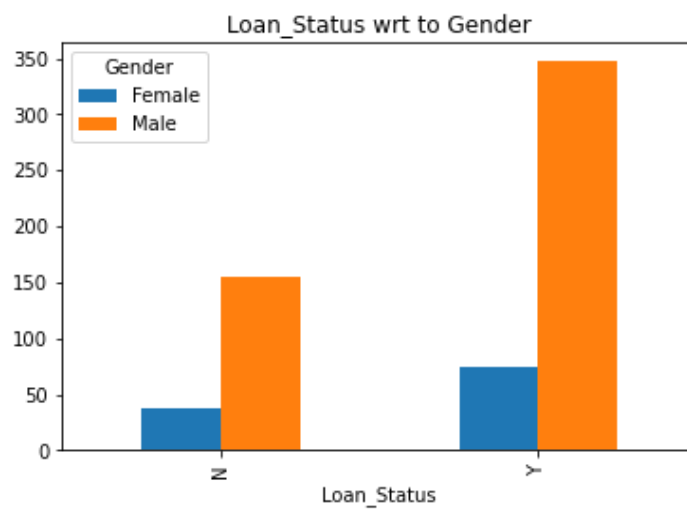


Figure 3.4

8. Also in further EDA for numerical attributes we create distplot visuals to get an idea about the density at each values.

9. Then we form a correlation plot with a heatmap to find out highly correlated variables in our dataset to eliminate redundancy. Also we remove unnecessary columns like loan_id which is not necessary for our training. Also we convert categorical data into numerical data with the help of label encoder to get better understanding.
10. Further for model building we specify input and output attributes , for loan approval analysis loan_status is an output variable.
11. Then we split the data set into training data and for test data. The first subset is used to fit the model and is referred to as the training dataset. The second subset is not used to train the model; instead, the input element of the dataset is provided to the model, then predictions are made and compared to the expected values. This second dataset is referred to as the test dataset.
12. To split our dataset we use train_test_split function from sklearn. Then with the help of scikit learn we use classification models like random forest, logistic regression etc. to find out the accuracy.
13. At last we create a confusion matrix . A confusion matrix, also known as an error matrix, is a summarized table used to assess the performance of a classification model. The number of correct and incorrect predictions are summarized with count values and broken down by each class.

		Actual Values	
		Yes	No
Predicted Values	Yes	True Positive	False Positive
	No	False Negative	True Negative

Figure 3.5 (Structure of Confusion matrix)

TP: Outcome where the model correctly predicts the positive class.

TN: Outcome where the model correctly predicts the negative class.

FP: Also called a type 1 error, an outcome where the model incorrectly predicts the positive class when it is actually negative.

FN: Also called a type 2 error, an outcome where the model incorrectly predicts the negative class when it is actually positive.

3.5 Insights :

Some of the insights which we get about our dataset after doing data analysis and visualization –

After doing Univariate analysis

- 80 % of the applicants are male in the training dataset.
- Nearly 70% are married.
- About 75% of loan applicants are graduates.
- 85% of applicants have a credit history of 1.

After doing Bivariate analysis

- Married applicants have slightly more chances of loan approval.
- Graduates have more chances of loan approval than non-graduates.
- It seems people with credit history as 1 are more likely to get their loan approved.
- The proportion of loan getting approved in the semi-urban areas is higher as compared to rural or urban areas.

CHAPTER 4

POWER BI

4.1 Introduction

Business intelligence (BI) is the set of techniques and tools for the transformation of raw data into meaningful and useful information for business analysis. To put it simply, Business intelligence is the technology which gets the right data to the right people, at the right time so that they can make more effective business decisions.

One of the main challenges faced by the clients in the financial sector is how to quickly and efficiently analyse large amounts of data and transform it into deep business insights that continuously drive the decision-making process to remain ahead of the competition.

In a nutshell, data visualization is nothing but the pictorial or graphical representation of information/ data. It provides insights into complex data sets by communicating the key aspects in more intuitive and meaningful ways.

4.1.1 Why data visualization is important?

The way, human brain processes information, it is easier to use images, charts or graphs to understand and to visualize large amounts of complex data, than to go through spreadsheets or reports. Take any image for example, we all know the phrase ‘An image is worth a thousand words’. This is completely true because images aren’t just a mere collection of pixels, they also hold a lot of information. This information in visual form is easy to understand than reading the same facts in text form.

Data visualization can help to:

- Identify key areas and hidden patterns.
- Get factors that give better customer insights.
- Make proper predictions.

4.2 What is Power BI ?

Power BI is a business analytics service provided by Microsoft. It provides interactive visualizations with self-service business intelligence capabilities, where end users can create reports and dashboards by themselves, without having to depend on information technology staff or database administrators.

In the financial services industry, time is literally money. Time spent chasing down data and creating static reports manually means less time for gaining insights, which leads to delayed responses to customers and rapidly shifting market conditions.

Power BI is redrawing this picture. Financial data teams no longer have to use up to 80% of their time trying to find, format, and present the right data, with little time left over to make the right decisions and partner fully with decision makers. Instead, Power BI enables analysts to share the latest, refreshed data to colleagues through powerful, interactive visualizations.

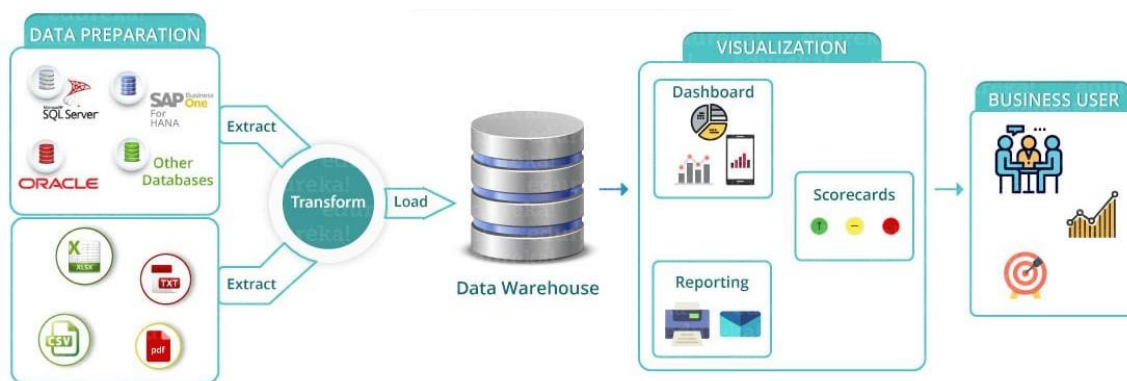


Figure 4.1 (POWER BI Architecture)

Power BI's architecture has three phases. The first two phases partially use ETL (Extract, Transform and Load) to handle the data. Let us take a look at these phases one by one:

1. Data Integration

An organisation can be required to deal with data that comes from different sources. The data from data sources can be in different file formats. The data is first extracted from different sources which can be your different servers or databases etc. This data is then integrated in a standard format and then stored at a common area called as staging area.

2. Data Processing

The integrated data is still not ready for visualization because the data needs processing before it can be presented. This data is pre-processed or cleaned. For example, missing values or redundant values are removed from the data set. After the data is cleaned, business rules are applied to the data and it is transformed into presentable data.

3. Data Presentation

So once the data is loaded and processed now it can be visualized much better with use of various visualizations that Power BI has to offer. Use of reports, dashboards help one represent data in more intuitive manner. These visuals, reports help business end users to take business decisions based on the insights.

4.3 Building Blocks of Power BI

Everything you do in Power BI can be broken down into following building blocks. The basic building blocks of Power BI are the following:

- Visualizations
- Datasets
- Reports
- Dashboards

Visualizations:

A visual representation of data is called visualization. For example, a chart or a graph can be used to represent data visually. Power BI gives you different visualization types, which keep getting updated with time. Following image shows some commonly used visualizations:

- Map representation
- Card visualization
- Stacked area chart
- Tree map
- Pie chart

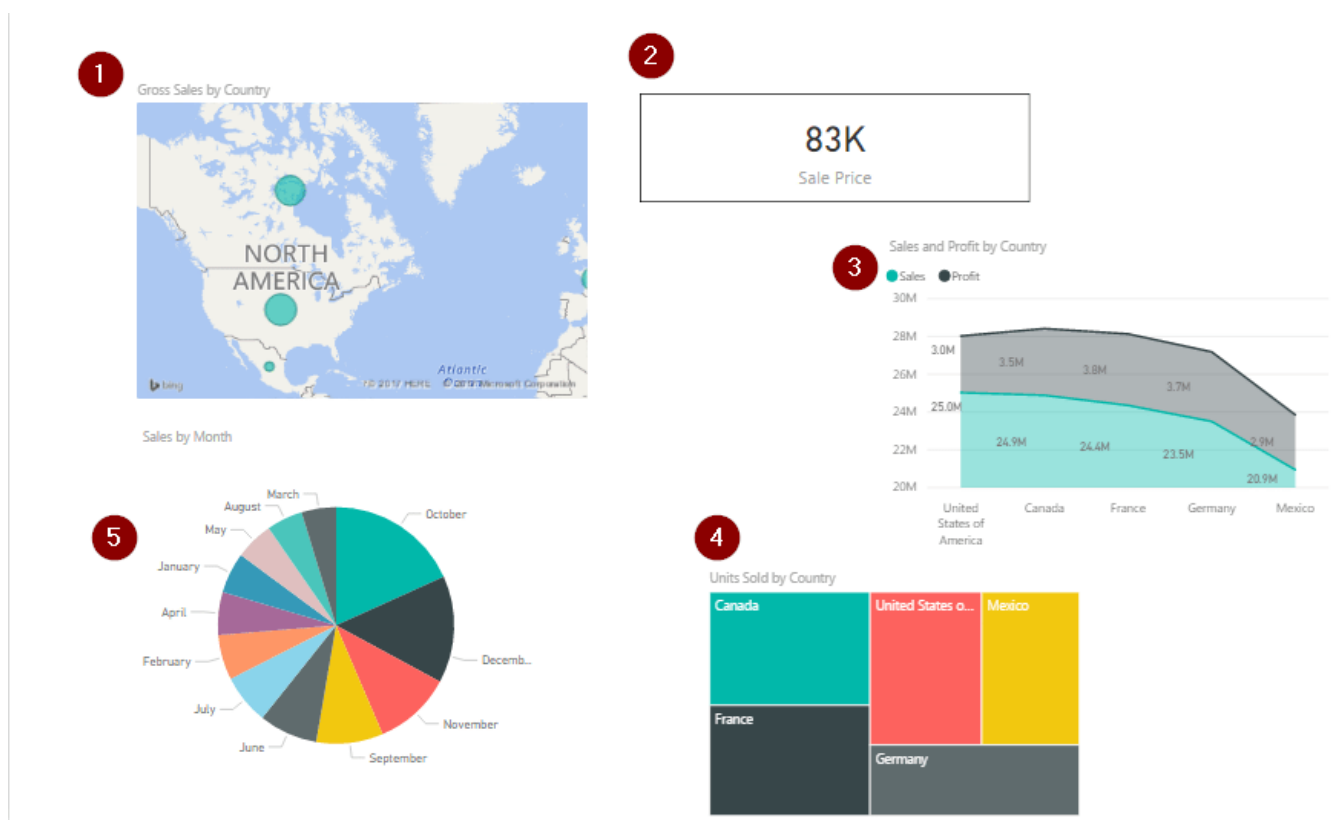


Figure 4.2

Datasets:

We know that data-set is nothing but a collection of data or information. Power BI harnesses this data to create visualizations. It can be a simple data set or a combination of many different sources, which can be filtered and combined to provide a different data set altogether.

For example, you can pull together data from many different sources like different database fields, an excel table, and online results of some email campaign to create the data set. Having said that, you may want to filter your data before you bring it into Power BI. Filtering lets you focus on the data that matters to you. The image below shows a sample data set.

Reports:

A collection of visualizations that appear together on one or more pages is a report in Power BI. It is a collection of items that are related to one another.

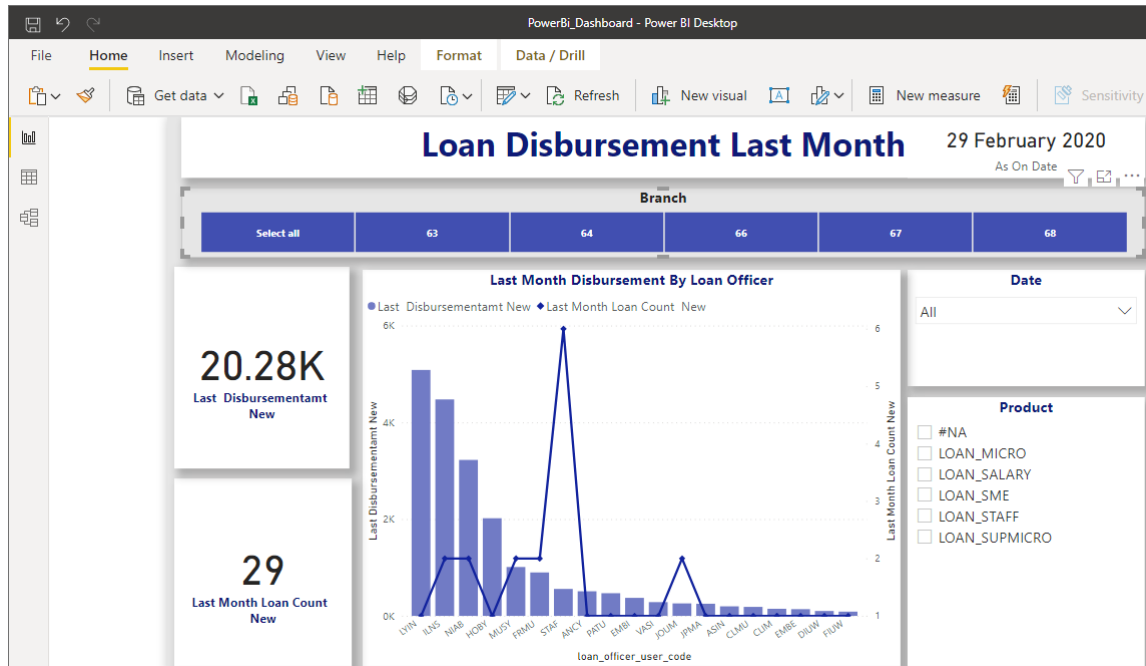


Figure 4.3

Dashboards:

A Power BI dashboard is a single page interface. It is often called a canvas, that uses visualizations to tell a story. Because it is limited to one page, a well-designed dashboard contains only the most-important elements of that story. The visualizations you see on the dashboard are called tiles and are pinned to the dashboard from reports.

In Power BI, a tile is a single visualization found in a report or on a dashboard. It's the rectangular box that contains each individual visual.

Power BI gives you the freedom to move or arrange tiles, so you can present the data the way you want to, even while you're creating a report or a dashboard. You can make the tiles bigger, change their height or width, and snuggle them up to other tiles the way you want.

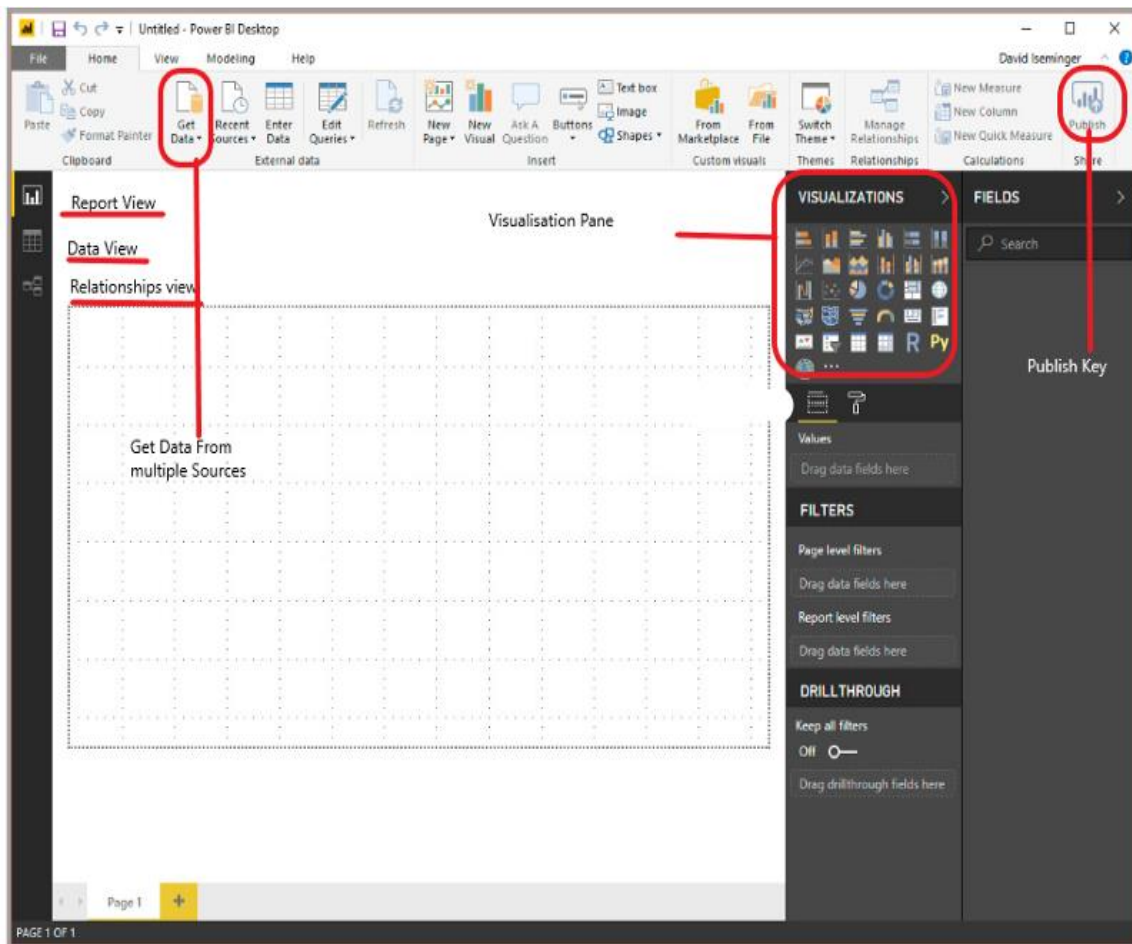


Figure 4.4(Power BI Interface Look)

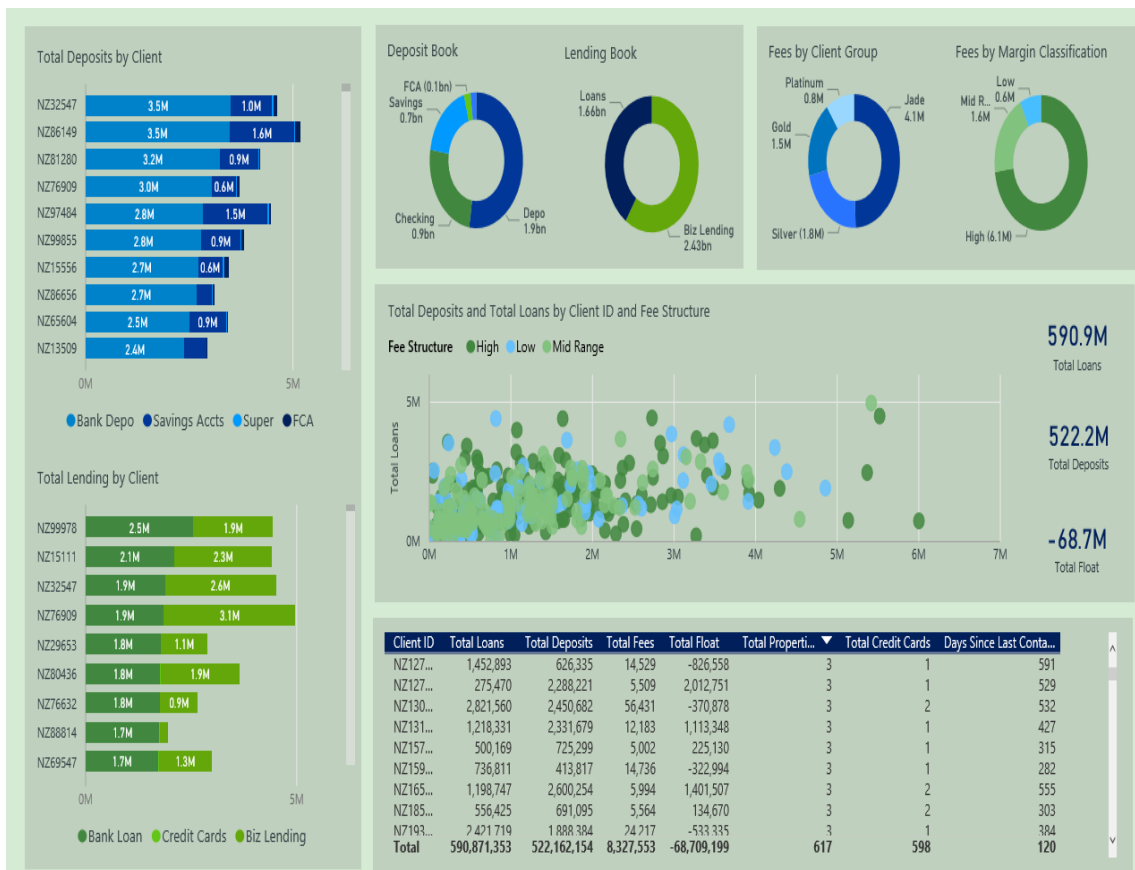


Figure 4.5(Dashboard view)

CHAPTER 5

LIMITATIONS

Disadvantages of Data Visualization

- It gives assessment not exactness.
- One of the biggest draws of visualization is its ability to take big swaths of data and simplify them to more basic, understandable terms. However, it's easy to go too far with this; trying to take millions of data points and confine their conclusions to a handful of pictorial representations could lead to unfounded conclusions, or completely neglect certain significant modifiers that could completely change the assumptions you walk away with.
- When users start relying on visuals to interpret data, which they can use at-a-glance, they could easily start over-relying on this mode of input.

Disadvantages of Power BI

- Power BI is good with handling simple relationships between tables in a data model. But, if there are complex relationships between tables, that is, if they have more than one links between tables, Power BI might not handle them well.
- The user interface of Power BI is often found crowded and bulky by the users. It is in the sense that there are many icons of options that block the view of dashboard or report.
- Power BI has a limit of ingesting data at a time which is approximately 2 GBs of data. If you wish to import and use data of even greater volumes, you need to extend your free version to a paid version of Power BI. Also, users have reported that Power BI takes a little more than usual time or even hangs while processing millions of rows and columns of data.

CHAPTER 6

FUTURE SCOPE

Business Analytics in India has remarkable prospects. A profession in Business Analytics is fulfilling and offers ample scope for learning and growth. Thorough knowledge of statistical techniques, quantitative capacity, business learning, logical thinking, Big Data, instruments to understand the accessible data, and asset management are some of the essential skills required to be a Business Analytics. Understanding the business situations and problem-solving abilities are other skills required.

Analytics can radically change the present business situation by capturing a large volume of data, expand business models, energize the imaginative procedures and overall growth and development of a company.

Analytics will play an important role in data security. Analytics are already transforming differential privacy, intrusion detection, digital watermarking and malware countermeasures. The Internet of Things (IOT) will continue to grow rapidly. Analytics tools and methods for dealing with large amounts of structured and unstructured data generated by IOT will continue to gain importance.

Companies will voice their need of routinely monetizing their own data for financial gain. Growth of Cognitive Analytics. Relevance of 'Open Source Solutions' will regain momentum. Boost in demand for Data Scientists- a hunt for people who can balance quantitative analysis skills with an ability to tell the story of their data in compelling, visual ways. Companies would become over-critical and cautious about Data Accuracy.

CHAPTER 7

CONCLUSION

The availability of Big Data, low-cost commodity hardware, and new information management and analytic software have produced a unique moment in the history of data analysis. During these age of big data, the business market and technology based industry has been totally radical and has caused both the professionals to continue to work together and deliver on promise. Data analytics is emerging as a field that is revolutionizing science and industries alike. Work across nearly all domains is becoming more data driven, affecting both the jobs that are available and the skills that are required. As more data and ways of analyzing them become available, more aspects of the economy, society, and daily life will become dependent on data for visualization and prediction of future trends for better profit margin and flourishing the businesses. In the loan prediction system, we saw how the raw data stored in database can be used fruitfully to analyze the customer trends and behavior to predict whether the consumer is good to get loan or not. In this way company will roughly have an idea, does the customer has the capability to repay back loan on time or not. These prediction saves company from millions of losses. This is the power of analysis of data. In the above work not only did i found out the eligibility of customer for loan but also other parameters such as- a. different region across the nation from where people are applying for loan, which can helped company to analyze, which states of India are performing well in their business and in which states do the company need to spread their business to generate more profit margin b. which are the common areas from where the customer are not able to repay back the loan etc. Also, we saw there are four types of data analysis namely- predictive, prescriptive, diagnostic and descriptive analysis. When all these are amassed, which includes applying of different machine learning models to generate the accuracy check, more better prediction can be made. All in all, business analytics and data analytics is a proliferating and very promising field. In India not all companies are very well aware of the these advanced technology, but some years from now, every business as well as technology oriented companies will be using this platform to give rise to their profit margin.

REFERENCES

- <https://www.javatpoint.com/power-bi>
- <https://www.upgrad.com/blog/data-preprocessing-in-machine-learning/>
- <https://data-flair.training/blogs/power-bi-advantages-and-disadvantages/>
- <https://www.guru99.com/what-is-data-analysis.html>