Data Analysis Portfolio

Prepared by:-
Rishi Vardhan Reddy Armoor

Professional Background

I am pursuing under graduation in Electronics and Communication Engineering and am in my final year with an 8.1 CGPA. Throughout my engineering, I acquired knowledge of the data domain and decided to work as a data analyst. I so started upskilling in data tools like Excel, Python, and SQL, and coursework like database management systems, machine learning, and visualization tool Tableau.

I did several certifications from LinkedIn and YouTube related to data analytics which helped to do my first internship at Techno Hacks as a Data Analyst and gain some real-time experience there. I also led a team of four members for my minor and major projects in the machine-learning field where we worked on Neural Networks, Exploratory Data analysis, and several libraries from Python.

I also published the document papers of my minor and major projects in an International research journal (IJR). It took me a while to decide on what domain I choose as my profession but it was all worth it because I enjoyed the process and looking forward to working as a data analyst in the real world and contributing to the company growth and my personal growth as well.

Table of Contents

Professional Description ---------- 2
Table of Contents ---------- 3
Data Analytics Process
- Project Description ---------- 6
- Insights ---------- 6
- Appendix ---------- 7

Instagram User Analytics
- Project Description ---------- 8
- Approach ---------- 8
- Tech-Stack Used ---------- 9
- Insights ---------- 9
- Conclusion ---------- 10
- Appendix ---------- 10

Operation & Metric Analytics
- Project Description ---------- 11
- Approach ---------- 12
- Tech-Stack Used ---------- 12
- Insights ---------- 12
- Conclusion ---------- 13
- Appendix ---------- 13

Hiring Process Analytics
Project Description ---------- 14
Approach ---------- 15
Tech-Stack Used ---------- 15
Insights ---------- 15
Conclusion ---------- 16
Appendix ---------- 16
IMDB Movie Analysis
Project Description ---------- 17
Approach ---------- 18
Tech-Stack Used ---------- 18
Insights ---------- 18
Conclusion ---------- 19
Appendix ---------- 20
Bank Loan Case Study
Project Description ---------- 21
Approach---------- 22
Tech-Stack Used ---------- 22
Insights ---------- 22
Conclusion ---------- 24
Appendix ---------- 24

Impact of Car Features
Project Description ---------- 25
Approach ---------- 26
Tech-Stack Used ---------- 27
Insights ---------- 27
Conclusion ---------- 28
Appendix ---------- 28
ABC Call Volume Trend
Project Description---------- 29
Approach ---------- 30
Tech-Stack Used ---------- 31
Insights ---------- 31
Conclusion ---------- 32
Appendix ---------- 32

Data Analytics Process:

Project Description:

We Use Data Analytics in everyday life without even knowing it.

For example: Going to the market to buy something.

Our task is to give examples of such a real-life situation where we can use Data Analytics and link it with the data analytics process. I need to prepare a ppt/PDF on a real-life scenario explaining it with the above process (plan, prepare, process, analyze, share, act) and submit it as part of this task.

Insights:

This is my first project in this course and it was helpful for me to understand the Applications in Real life Scenario.

Appendix:

Link to the project (ctrl + tap)

Instagram User Analytics:

Project Description:

The Project is about the analytics of users' Instagram accounts. This project aims to understand the ways to provide a better experience and also a few intimations to the user based on their account use. I first understand the data provided in the commands then break down the problems into small parts and try to solve them.

Approach:

The project was a bit challenging because it was my first project in SQL but I tried to solve it by breaking down the problem which made me understand the data and the questions easily. I executed the project by creating a database named 'ig_clone' and created tables as per the commands provided and the inserted values into tables. There were a few commands I didn't understand like TIMESTAMP.

Tech-Stack Used:

I used SQL Server Management Studio. The reason I chose this stack is that I was familiar with this studio since I practiced SQL earlier in college.

Insights:

I gained a lot of knowledge while doing this project. I discovered new commands in SQL and also gained experience in how to handle real-world data and how to approach the requirements of investors.

i) The significant insight related to the project is, that there are very few users on Instagram who didn't post a single photo on Instagram.

ii) Also most people register on Instagram on the first and second days of the week which are the weekdays surprisingly and at least on the weekend which is on the seventh day of the week.

Conclusion:

I completed the project in three days. This is an achievement for me because I initially felt that I might not complete the project or take a longer time than I completed it. This Project built confidence in me by completing the project. I also understood the analytics of the user's usage on Instagram. Though I use Instagram daily, after completing this project I understood how the user's usage works.

Appendix:

Link to the project (ctrl + tap)

Operation & metric Analytics:

Project Description:

Operational Analytics is a crucial process that involves analyzing a company's end-to-end operations. This analysis helps identify areas for improvement within the company. As a Data Analyst, you'll work closely with various teams, such as operations, support, and marketing, helping them derive valuable insights from the data they collect.

One of the key aspects of Operational Analytics is investigating metric spikes. This involves understanding and explaining sudden changes in key metrics, such as a dip in daily user engagement or a drop in sales. As a Data Analyst, you'll need to answer these questions daily, making it crucial to understand how to investigate these metric spikes.

our goal is to use your advanced SQL skills to analyze the data and provide valuable insights that can help improve the company's operations and understand sudden changes in key metrics.

Approach:

I take this project as a new challenge and learn new things from this project. As I mentioned earlier, I executed the analysis by taking the required tables and after successfully getting the output studying it until I understand it and then coming to a conclusion about the analysis.

Tech-Stack Used:

I used ‘Microsoft SQL Server Management Studio’ in this project. I Used this server because I am familiar with this software and easy to type and execute the code.

Insights:

I learned about the timestamp and datetime formats from this project which was quite challenging because this was the first time I was using.

The key observations I made through this project are:

i) The Persian language has the highest share percentage among all other languages.

ii) The email engagement is around 90,000 which was huge. That means the major mode of communication is through mail.

iii) The Macbook Pro was the second in the weekly user engagement per device. There are no other users related to Android near Mac.

Conclusion:

I got to know about new topics related to SQL. Timestamp and DateTime functions. Also, I understood how to import the data from a CSV file into the management studio.

Appendix:

Link to the Project (ctrl + tap)

Hiring Process Analytics:

Project Description:

Imagine you're a data analyst at a multinational company like Google. Your task is to analyze the company's hiring process data and draw meaningful insights from it. The hiring process is a crucial function of any company, and understanding trends such as the number of rejections, interviews, job types, and vacancies can provide valuable insights for the hiring department.

As a data analyst, I'll be given a dataset containing records of previous hires. Your job is to analyze this data and answer certain questions that can help the company improve its hiring process. The goal of this project is to use your knowledge of statistics and Excel to draw meaningful conclusions about the company's hiring process. Your insights could potentially help the company improve its hiring process and make better hiring decisions in the future.

Approach:

After making a copy of the dataset I started solving it one by one. I first created the required columns and applied the Excel formulas to calculate the minimum salary, maximum salary, and average. But to be specific it was challenging to create the class intervals.

Tech-Stack Used:

I used Microsoft EXCEL 2021 Version which is inbuilt on my computer. The purpose of using EXCEL is because the task demands to use of this software and it's also easy in EXCEL compared to other software for data analysis as it is a direct add-in function.

Insights:

The major insights I draw from this project is that the number of employees hired in the company are mostly men and the gap between the minimum salary and maximum salary is huge. The class intervals I took were 12 as I felt it was required for the maximum salary. The knowledge I

Conclusion:

I was unfamiliar with EXCEL before the project but after completing the Project I got used to Excel formulas and learned how to create bar graphs, histograms, interval width, and class intervals.

Appendix:

Link to the project (ctrl + tap)

IMDB Movie Analysis:

Project Description:

The dataset provided is related to IMDB Movies. A potential problem to investigate could be: "What factors influence the success of a movie on IMDB?" Here, success can be defined by high IMDB ratings. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

Data Cleaning: This step involves preprocessing the data to make it suitable for analysis. It includes handling missing values, removing duplicates, converting data types if necessary, and possibly feature engineering.

Data Analysis: Here, you'll explore the data to understand the relationships between different variables. You might look at the correlation between movie ratings and other factors like genre, director, budget.

Approach:

The approach is simple, just understand the data and remove the duplicates if any. Then, thoroughly read the description and hints provided to solve the queries. Also, this project is a bit lengthy so I divided the problems into pieces and solved them so that won't be confusing.

Tech-Stack Used:

I used Microsoft EXCEL 2021 Version which is inbuilt on my computer. The purpose of using EXCEL is because the task was demanding to use this software and it's also easy in EXCEL compared to other software for descriptive analysis as it is a direct add-in function. Also, the data set provided was in an Excel sheet so it's easy to proceed with it.

Insights:

I understood how to calculate the profit margin for film and also got to know about variance and standard deviation in real-life examples.

Understood descriptive statistics, histograms, and how to depict the relation between any two columns I Also, gained knowledge on how to convert a single column into multiple ones and apply filters too.

The key finding is that movies were released in mostly “English” other than English there were very few movies from other languages. Also, most movies were released in the “Drama” genre. There are no movies with more than 9 Average IMDB rating.

Result:

I understand more clearly the statistics and their characteristics in real-world data. Like taking the correlation, and percentile to find out the range of input. I also got to know how to hyperlink the Excel sheet into the ppt and improved my communication skills by doing a 5-minute video on this project. The mindset of not only understanding the data and the queries but also understanding the patterns and domain knowledge contributed in a great way while doing this project.

Appendix:

Link to the project (ctrl + tap)

Bank Loan Case Study:

Project Description:

Imagine you're a data analyst at a finance company that specializes in lending various types of loans to urban customers. Your company faces a challenge: some customers who don't have a sufficient credit history take advantage of this and default on their loans. Your task is to use Exploratory Data Analysis (EDA) to analyze patterns in the data and ensure that capable applicants are not rejected.

The main aim of this project is to identify patterns that indicate if a customer will have difficulty paying their installments. This information can be used to make decisions such as denying the loan, reducing the amount of the loan, or lending at a higher interest rate to risky applicants. The company wants to understand the key factors behind loan default so it can make better decisions about loan approval.

Approach:

To gain some domain knowledge I started reading about Risk Analytics in banking and financial services. Then, started breaking each assigned task into bits and pieces so that the problems could be solved in a step-by-step manner.

I made sure that I got the domain knowledge first and read the given Excel sheets thoroughly to understand every column and then start solving the queries.

Tech-Stack Used:

I used the Microsoft Excel 2021 version which was available on my computer and also this project requires Microsoft Excel to solve the tasks.

Insights:

The first insight was the clients with payment difficulties were much lower than I thought. There were only 4026 people out of 50,000 who faced difficulty regarding payments.

Secondly, there were outliers which was a problem while understanding the data because that could disturb the insights. Also while handling the missing data I found a lot of missing values and deleted most of the columns which were not important for the case study.

The Interesting Insight for me was while doing the univariate analysis I found that the people applying for loans were mostly Female. There were almost double the Female members as clients for the bank than men.

When comparing the amount of income vs the status of the payments paid. There was no big difference between the incomes of people who paid the loans on time and people who faced diffusivities in payments. But mostly there are people in between the range of 0 to 1000000. More “Working” people applied for the loan then came “Commercial associates” followed by “Pensioners” and “State servants”. Regarding the Target, state servants were paying the payments without any issues, and then came pensioners.

Result:

I gained a lot of knowledge in banking and financial domains. In the tech stack, I gained more experience in Excel. Especially In Univariate, Bivariate, and segmented analysis.

I also understood that understanding the data is very important rather than directly jumping into the problems. Also, when handling real-world data we need to invest more time in handling the missing data and outliers in every column.

Appendix:

Link to the Excel sheet (ctrl + tap)

Impact of Car Features:

Project Description:

The automotive industry has been rapidly evolving over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation. With increasing competition among manufacturers and a changing consumer landscape, it has become more important than ever to understand the factors that drive consumer demand for cars.

In recent years, there has been a growing trend towards electric and hybrid vehicles and increased interest in alternative fuel sources such as hydrogen and natural gas. At the same time, traditional gasoline-powered cars remain dominant in the market, with varying fuel types and grades available to consumers.

For the given dataset, as a Data Analyst, the client has asked How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand?

This problem could be approached by analyzing the relationship between a car's features, market category, and pricing, and identifying which features and categories are most popular among consumers and most profitable for the manufacturer. By using data analysis techniques such as regression analysis and market segmentation, the manufacturer could develop a pricing strategy that balances consumer demand with profitability, and identify which product features to focus on in future product development efforts. This could help the manufacturer improve its competitiveness in the market and increase its profitability over time.

Approach:

The Approach for this project is simple, as it is an end-to-end project we are required to understand the case scenario of the project make our queries, and decide on what columns will meet the better understanding of any respective scenario.

Also, It is essential to gain some domain knowledge. So firstly I will research this domain and then study the data set provided and then start doing the spreadsheet. Then break down the project into bits and solve them separately to avoid confusion.

Tech-Stack Used:

The Tech stack used in this project is Microsoft Excel 2021 Version. We can also use Python for data analysis but in Excel the market segmentation is easily performed and can be easily understood.

Insights:

There are many insights drawn from this project. But I would like to mention three insights that stand out among all they are,

(i) The maker of the car, “Maybach” has the highest average of the Manufacturer’s Suggested Retailed Price(MSRP). But surprisingly the Average popularity of the company is very low i.e. just 67 where the highest popularity is 3916 for “Audi” but again the MSRP is low i.e. just 61,457 dollars.

(ii) While looking at the pivot table where every car maker’s popularity, MSRO, and Engine HP are compared unlike the above insight, “Ford” has the highest Popularity with the sum of 49,83,817, and as expected the MSRP is low and engine HP is decent.

(iii) The Correlation between ‘MSRP’ and ‘Engine HP’, and ‘Engine Cylinders’ is positive, and ‘Highway MPG’, and ‘City MPG’ is a negative correlation.

Conclusion:

This Project helped me understand the car manufacturing domain and also how to handle the data if there are no queries given and need to do it on my own.

When it comes to tech stack, it’s always a learning process as we get to learn new concepts in every project. So in this project, I got to understand the “Regression Analysis” and “Market segmentation” in Excel. Which was challenging but helped to understand the patterns and trends in this project.

Appendix:

Link to Excel sheet (ctrl + tap)

ABC Call Volume Trend:

Project Description:

In this project, you'll be diving into the world of Customer Experience (CX) analytics, specifically focusing on the inbound calling team of a company. You'll be provided with a dataset that spans 23 days and includes various details such as the agent's name and ID, the queue time (how long a customer had to wait before connecting with an agent), the time of the call, the duration of the call, and the call status (whether it was abandoned, answered, or transferred).

A Customer Experience (CX) team plays a crucial role in a company. They analyze customer feedback and data, derive insights from it, and share these insights with the rest of the organization. This team is responsible for a wide range of tasks, including managing customer experience programs, handling internal communications, mapping customer journeys, and managing customer data, among others.

In the current era, several AI-powered tools are being used to enhance customer experience. These include Interactive Voice Response (IVR), Robotic Process Automation (RPA), Predictive Analytics, and Intelligent Routing.

One of the key roles in a CX team is that of the customer service representative, also known as a call center agent. These agents handle various types of support, including email, inbound, outbound, and social media support.

Inbound customer support, which is the focus of this project, involves handling incoming calls from existing or prospective customers. The goal is to attract, engage, and delight customers, turning them into loyal advocates for the business.

Approach:

I first spent a lot of time, understanding the Business domain and thoroughly knowing each column from the dataset, and then breaking down the required columns for respective columns which can avoid confusion in solving the given questions.

Instead of applying all the Excel formulas all at a time, I inserted new columns applied the newly inserted columns to the next column, and solved queries.

Tech-Stack Used:

I used the Microsoft Excel 2021 version in this project because the project demanded providing a spreadsheet that can be done in Excel and the version was built on my computer.

Also, I was very much aware of Excel and its properties which made my work a bit easy.

Insights:

By doing this project I gained a lot of knowledge, especially in performing the Excel functions which included the mathematical formulas in it. I also got to know about Customer Experience Analytics and business understanding from the link provided in the project description. Now I am well aware of the CX analytics.

Also understood how to get the data based on the assumptions made and implement that data into the queries to get the output as shown in night shift manpower planning.

Conclusion:

I conclude that this project made my Excel skills even stronger and also got to know about a new domain called Customer Experience (CX) analytics. This domain understanding helped to learn more about the ABC Call Volume analysis.

The queries in this project were not directly solved but depended on various factors that we needed to calculate using mathematical functions.

Appendix:

Link to Excel Sheet (ctrl + tap)