DATA ANALYSIS

PORTFOLIO

By :

Rachit Dani

rachitdani

rachitdani2014@gmail.com

About Me

I have just completed my BTech in Computer Science in 2023 from the University of Mumbai with an aggregate of 8.74 CGPA.

I have lately completed my Final Year Project on the topic "Commercial Band Detection with details" and this project has been given a Letter of Appreciation from Zee Television Channel.

Data is the thing that fascinates me based on data we can solve a lot of problems.
I possess technical skills in Python, SQL, R Language, Machine Learning and Tableau.
Some Soft Skills are Story Telling, Teamwork, Presentation and Analytical Thinking.

I am a dedicated learner looking forward to gaining experience in the industry

Rachit Dani

Data Analyst / Data Scientist

rachitdani2015@gmail.com
8667279701
Banglore, India
linkedin.com/in/rachitdani

EDUCATION

Bachelor Of Engineering In Computer Science
University Of Mumbai
06/2019 - Present *8.74 CGPA*

Central Board Of Secondary Education
CS ACADAMY, Coimbatore
06/2017 - 05/2019 *80.00%*

Indian Certificate Of Secondary Education
HIHS, Coonoor
07/2015 - 05/2017 *85.33%*

PROJECTS

Commercial Band Detection With Details (07/2022 - 04/2023)
- In This project we take the input in the .mp4 format and then break it into images and then use Yolov7 to detect and crop the L and I shaped bands and then extract the text as output from the bands by using G-Vision OCR.

Bank Loan Case Study (03/2023 - 04/2023)
- Using Exploratory Data Analysis to analyze patterns in the present data and to ensure that the applicants capable of paying the loan are not rejected

IMDB Movie Analysis (03/2023 - 04/2023)
- This project comprised the use of pivot tables along with the statistical mathematical concepts and then visualizing them.

Hiring Process Analytics (02/2023 - 03/2023)
- This project used excel concepts right from cleaning the data and using VLOOKUP function to the use of pivot tables

Instagram User Analytics , Operation Analytics and Investigating Metric Spike (01/2023 - 02/2023)
- This project had queries which had basic SQL and advanced SQL concepts like window functions ,nested queries and so on.

BigMart Sales Predictor (01/2022 - 05/2022)
- By using the dataset if the sales data we have compared various algorithms for getting the maximum accuracy

IMAGO : Traffic Sign Detector (07/2021 - 12/2021)
- A flask based ui website where the user inputs the traffic sign as an image and we display the output of the sign using ML models.

SKILLS

Python(numpy, pandas, scikit-learn) | SQL | Microsoft Excel (VBA) | Machine Learning Algorithms | R Language(tidyverse, janitor, skimr , dplyr, R markdown) | Tableau

ACHIEVEMENTS

Letter Of Appreciation From ZEE Television Channel (08/2022 - 04/2023)
Recieved for the project Commercial Band Detection With Details

CERTIFICATES

Google Data Analytics Specialization Course (11/2022 - 04/2023)
An in-detail data analytics course where I learnt the process of data collection, cleaning, analyzingand then creating reports of the process

Trainity Data Analytics (01/2023 - Present)
A in-detail project-based course where I learnt the advanced concepts of sql (window functions, nested queries) excel , statistics, and visualizations.

LANGUAGES

English
Full Professional Proficiency

Hindi
Full Professional Proficiency

Gujarathi
Full Professional Proficiency

TABLE OF CONTENTS

Data Analysis Process

Example : Say we are going out on a Sunday with Friends

Plan

First, we make a plan about which place will we go. After asking everyone where will we go out we will decide on a place say Rcity Mall, Mumbai.

Prepare

Now we decide how much money we have in our pocket for spending.

Process

Now we decide what to do in the mall like shopping or eating food or going bowling

Analyze

You finally decided to eat so now you decide what to eat based on the amount of hunger you have. Since you were moderately hungry you go for a Grilled Sandwich

Share

Now you tell the restaurant biller that you want a Grilled Sandwich

Act

Now you finally buy it

Instagram User Analytics

Project Description

User analysis is the process by which we track how users engage and interact with our digital product (software or mobile application) in an attempt to derive business insights for marketing, product & development teams.
These insights are then used by teams across the business to launch a new marketing campaign, decide on features to build for an app, track the success of the app by measuring user engagement and improve the experience altogether while helping the business grow.

Approach

1. Create a Database : Use SQL commands to create a database and tables.
2. Perform Analysis : Run SQL command to perform analysis.
3. Create a Report : Finally Create a report documenting the entire process.

Findings

1. **Marketing :**
    a. Rewarding Most Loyal Users
    b. Remind Inactive Users to Start Posting
    c. Declaring Contest Winner
    d. Hashtag Researching
    e. Launch AD Campaign
2. I**nvestor Metrics :**
    a. User Engagement
    b. Bots & Fake Accounts

Tech Stack Used

Coding Language : MySQL
Platform Used : db Fiddle

Result

On completing this project I have learned when and where to use JOINS, SUBQUERIES , AGGREGATOR FUNCTIONS, GROUPBY and ORDERBY commands

Drive Link

https://drive.google.com/file/d/14kC6aWNrT0PucpLGk2JBWkfCmJsT7R67/view

Operation and Metric Analysis

Project Description

Operation Analytics is the analysis done for the complete end-to-end operations of a company. With the help of this, the company then finds the areas on which it must improve upon. You work closely with the ops team, support team, marketing team, etc and help them derive insights out of the data they collect.
Investigating metric spikes is also an important part of operation analytics as being a Data Analyst you must be able to understand or make other teams understand questions like- Why is there a dip in daily engagement? Why have sales taken a dip? Etc. Questions like these must be answered daily and for that its very important to investigate metric spikes.

Approach

Create a Database & Tables : Use SQL commands to create a database and tables.
Perform Analysis : Run SQL command to perform analysis.
Create a Report : Finally Create a report documenting the entire process.

Findings

- **User Engagement**
- **User Growth**
- **Weekly Retention**
- **Weekly Engagement**
- **Email Engagement**

Tech Stack Used

Coding Language : MySQL
Platform Used : Mode.com

Result

On Completing this project i had a clear understanding of JOINS,Nested Queries,EXTRACT ,ROUND and various other functions

Drive Link

https://drive.google.com/file/d/1kA-1mWwQvbSSwb_NcTHET-0W-MKviHAE/view

Hiring Process Analytics

Project Description

Hiring process is the fundamental and the most important function of a company. Here, the MNCs get to know about the major underlying trends about the hiring process. Trends such as- number of rejections, number of interviews, types of jobs, vacancies etc. are important for a company to analyse before hiring freshers or any other individual.

Approach

Clean the Dataset : Use Exploratory Data Analysis to clean the dataset
Perform Analysis : Use Basic Excel Functions and Charts to Visualize
Create a Report : Finally Create a report documenting the entire process.

Findings

- **Hiring**
- **Average Salary**
- **Class Intervals**
- **Charts and Plots**

Tech Stack Used

Platform Used : Google Sheets

Result

On Completing this project I had a clear understanding of data cleaning in Excel, Excels Conditional Filtering,Excel Functions and Charts

Drive Link

https://drive.google.com/file/d/1oEoEUGKNNYHB2Is9htooZ-r7UWwzVcAk/view

IMDB Movie Analysis

Project Description

The dataset contains various columns of different IMDB Movies. We must analyze the data and provide insights on the findings of the project

Approach

Clean the Dataset : Use EDA to clean the dataset
Perform Analysis : Use Excel Functions ,Pivot Tables and Charts to Visualize
Create a Report : Finally Create a report documenting the entire process.

Findings

- **Cleaning the data**
- **Movies with highest profit**
- **Top 250**
- **Best Directors**
- **Popular Genres**
- **Charts**

Tech Stack Used

Platform Used : Google Sheets

Result

On Completing this project I had a clear understanding of Data Cleaning in Excel,Various Excel Functions ,Pivot tables and Charts

Drive Link

https://drive.google.com/file/d/1bDgYRrim3san499Y2nlWBFth1JS8uzrb/view

Bank Loan Case Study

Project Description

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customers. You have to use EDA to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

Approach

Clean the Dataset : Use EDA to clean the dataset
Perform Analysis : Use python code to create charts
Create a Report : Finally Create a report documenting the entire process.

Findings

- **Using EDA to clean the dataset**
- **Identify and remove outliers**
- **Perform Segmented univariate, univariate and bivariate analysis**
- **Correlation between variables**
- **Summarize and Visualize Findings**

Tech Stack Used

Platform Used : Google Collab
Coding Language : Python

Result

On Completing this project I had a clear understanding of linking google drive to collab, data cleaning in python, finding correlation between variables and creating charts in python

Drive Link

https://drive.google.com/drive/folders/1Yf6gTBfyq0yHCEIFiYaTGGUtlll8LYP9

Analyzing the Impact of Car Features on Price and Profitability

Project Description

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

Collect the Dataset : "Car Features and MSRP" dataset from kaggle
Clean the Dataset : Use EDA to clean the dataset
Perform Analysis : Use python code to create charts
Create a Report : Finally Create a report documenting the entire process.

Findings

- **How does the popularity of a car model vary across different market categories?**
- **What is the relationship between a car's engine power and its price?**
- **Which car features are most important in determining a car's price?**
- **How does the average price of a car vary across different manufacturers?**
- **What is the relationship between fuel efficiency and the number of cylinders in a car's engine?**

Tech Stack Used

Platform Used : Google Collab
Coding Language : Python

Result

On Completing this project I had a clear understanding of linking google drive to collab, data cleaning in python, finding correlation between variables and creating charts in python

Drive Link

https://drive.google.com/drive/folders/1Yf6gTBfyq0yHCEIFiYaTGGUtlll8LYP9

ABC Call Volume Trend Analysis

Project Description

We are provided with a dataset of a Customer Experience (CX) Inbound calling team for 23 days containing various columns.
Data includes Agent_Name, Agent_ID, Queue_Time [duration for which customer have to wait before they get connected to an agent], Time [time at which call was made by customer in a day], Time_Bucket [for easiness we have also provided you with the time bucket], Duration [duration for which a customer and executives are on call], Call_Seconds [for simplicity we have also converted those time into seconds], call status (Abandon, answered, transferred).

Approach

Collect the Dataset : Provided by the CX Inbound Calling Team
Clean the Dataset : Use EDA to clean the dataset
Perform Analysis : Use Excel Functions,Pivot Tables and charts to visualize.
Create a Report : Finally Create a report documenting the entire process.

Findings

- **Calculate the average call time duration for all incoming calls received by agents in each Time Bucket.**
- **Show the total volume/ number of calls coming in via charts [Number of Calls vs Time graph]**
- **Calculate the Abandon Rate**
- **Propose a manpower plan required during each time bucket [between 9am to 9pm] to reduce the abandon rate to 10%.**
- **Propose a manpower plan required during each time bucket [between 9pm to 9am]. Maximum abandon rate is 10%.**

Tech Stack Used

Platform Used : Microsoft Excel

Result

On Completing this project I had a clear understanding Excel Functions,Pivot tables,Charts using problem-solving skills based on the data insights obtained and provide solution

Drive Link

https://drive.google.com/drive/folders/1vB7NKy7YG1h8JAy1veIW3Iyp77t7rZRD?usp=sharing

Conclusion

The Learnings I learnt from the above Projects are :

- The Stages of Data Analytics Process
- Application of Basic and Advanced SQL concepts in real-world scenarios.
- Importance of Data Cleaning and Outlier Detection as it has a very strong impact on our analysis.
- Application of Microsoft Excel features for data cleaning and visualizing.
- Using Python code to clean and analyze large datasets.
- Using Google Collab for reports as well as processing and visualizing the data
- Analytical Thinking as a data analyst to solve real-world problems
- Ability to apply all the learned concepts in real-world scenarios.