**DATA ANALYSIS PORTFOLIO:**

**PREPARED BY: MANPREET KOUR**

**Professional Background:**

Currently in my first year pursuing M.Sc. Chemistry .I have secured 8.82 CGPA (till 1st semester) and have several skills including Data Analysis, Python, SQL, Google cloud Program, Power BI, Tableau, Excel. I have worked with several companies as an intern like Sparks Foundation, Trainity ,Learn and build etc, as a Data analytic enthusiast .I am also working as a Business Analyst with a electric vehicle startup “MYeKEGAI” where I have worked on their Business Analytics project from scratch and managed different teams. I have worked on several projects related to data analysis, business analysis. As I am a fresher it would be great to experience the real challenges of the corporate world and understand how things work. Being a fresher, I think I am very flexible and adaptive to learn new things. I have theoretical knowledge. But I am waiting to use my theoretical knowledge in a practical way. And I believe by putting significant efforts I will learn.

**Table of Contents:**

1. Module-1 Project
2. Module-2 Project
3. Module-3 Project
4. Module-4 Project
5. Module-5 Project
6. Module-6 Project
7. Module-7 Project
8. Module-8 Project
9. Conclusion

**MINOR PROJECT 1:**

**Data Analytics Process**

The task is to give the example(s) of such a real-life situation where we use Data Analytics and link it with the data analytics process . Here I have taken two persons Mr Tim and Miss Susan and decided which superstore they should go to buy items so it will be cost effective.

LINK: <https://docs.google.com/spreadsheets/d/1H1nSt7spJ-KIVZ1m9cQR-uoL_RNvzm7BMcg8FqD7NOE/edit?usp=drive_link>

**MINOR PROJECT 2:**

Instagram User Analytics

**Project Description:**

Here we have done Instagram data analysis. Done all the task

mentioned by the authorities. The tasks gave us some beneficial

insights. 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:**

Performed all of the task given using SQL. Created all the tables

and data base as mentioned in the dataset provided.

**Tech-Stack Used:**

My sql server only

**Insights:**

Wider approach of what we are supposed to do in

Industry .Getting prepared for the actual work.

Result: We got many insights in this project. 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.

LINK: <https://drive.google.com/file/d/15FeFTcH21iBlasMAI-HnSzj2J2V_qDm9/view?usp=sharing>

**MINOR PROJECT 3:**

**Operation & Metric Analytics**

**Project Description:**

Here we have done Operational metric  analysis. Done all the task mentioned by the authorities. 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 spike 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.

**Approach:**

Performed all of the task given using SQL .Created all the tables and data base as mentioned in the dataset provided.

**Tech-Stack Used:** Google cloud program/Big query

**Insights:** Wider approach of what we are supposed to do in industry. Getting prepared for the actual work.

**Result:**  We got many insights in this project. 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.

LINK: <https://docs.google.com/document/d/1oHmHadybAm_FwFyAq5ibpDNOcqkXAMmR/edit?usp=sharing&ouid=112641900185186889653&rtpof=true&sd=true>

**MINOR PROJECT 4:**

**Hiring Process Analytics**

###### **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. Here we have analysed the male and female working population. People rejected and hired on basis of gender. Average salary given by our company People working in different department to analyse where work force is required where it is not.

LINK:

<https://1drv.ms/x/s!AgVOadNMXixUoW5_K8uXokmG3il7?e=2wV7Eq>

**MAJOR PROJECT 1:**

**IMDB Movie Analysis**

**There is a** dataset having various columns of different IMDB Movies. I have done the below mentioned tasks

1. **Movies with highest profit**

**Task:** Find the movies with the highest profit?

1. **Top 250:**

**Task:**Find IMDB Top 250

1. **Best Directors:**   
    **Task:**Find the best directors
2. **Popular Genres:**Perform this step using the knowledge gained while performing previous steps.  
   **Task:**Find popular genres

**LINK:** <https://1drv.ms/x/s!AgVOadNMXixUoXE5syHYINLkfwcS?e=vOW7k9>

**MAJOR PROJECT 2:**

**BANK LOAN CASE STUDY:**

**PROJECT DESCRIPTION**: Here we are performing EDA and also gaining basic grasp of risk analytics in banking and financial services, as well as how data is utilized to reduce the risk of losing money when lending to consumers.

We are using two datasets Application and previous application .Not going to use column description as its just tell about the information of variable used in other datasets.

**TECH STACK USED:** Power BI and Microsoft excel

**STATISTICAL EVALUATION** : We have taken out mean, median third quartile ,first quartile , Inter quartile range, inner outlier range, outer outlier range.

Approach

First started by cleaning both datasets and removing blank cells and unrequired columns.

For that I have used count blank function in each row and removed that column with excess of blank cells.

Data cleaning

All the values outside the inter quartile range and outer quartile range were removed ,this is called as removing outliers.

As this is cleaned dataset columns with Count blank equal to zero or less than 5% is there .Rest all the columns with high values of count blank were removed.

All the numerical values have there statistical evaluation by mean,median,quartiles etc but this is not followed for categorical values

categorical values-(non-numerical variables)- person's occupation, education status.

Numerical variables - income, credit etc.

Extracted the dataset parts with columns we are going to use and made a box whisker plot of same.

LINK:

|  |  |  |
| --- | --- | --- |
| https://docs.google.com/spreadsheets/d/1re2qqj-U2EttUpc61CiWt5WPuksmMAOd/edit?usp=sharing&ouid=112641900185186889653&rtpof=true&sd=true |  |  |

**MAJOR PROJECT 3:**

**Impact of Car Features**

Project description:

* Project is all about how can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand.
* This problem could be approached by analysing the relationship between a car's features, market category, and pricing.

Approach:

* The dataset contains information on various car models and their specifications, and is titled "Car Features and MSRP".
* First, I have cleaned the data and then used co-relation and regression analysis to find relation between all variables.

Tech stack used:

* Power BI
* Microsoft excel.

Data Cleaning:

I have cleaned the car data by removing the duplicates from the remove duplicate section of data tab and then removed all Blank values from find and select option in ribbon from the home button in tool bar.

Co-relation:

 Correlation analysis in research is a statistical method used to measure the strength of the linear relationship between two variables and compute their association.

INFERENCES:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| * The R squared value of Engine horsepower is maximum. It proves that out of all the variables "Engine HP" | | | | | |
| has maximum dependency (42%) on MSRP.   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | * Moreover, the r value for engine HP is 0.6 (near to 1) so, strong positive co- relation is there.  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | * As, seen from graph the dataset has negative slope and positive intercept and | | | | | | | | * correlation co-effecient is negative means there is negative co-relation and | | | | | | | | * increase in one variable lead to decrease in other. | | | |  |  |  | |  |  |  |  |  |  |  | | * As, value of r (pearson coeffecient) is near to -0.61471 ,that shows strong | | | | | | | |  |  | |  |  |  |  |  |

**LINK:** [**https://1drv.ms/x/s!AgVOadNMXixUoUWTwymL8-ZoAaxd?e=rycr2P**](https://1drv.ms/x/s!AgVOadNMXixUoUWTwymL8-ZoAaxd?e=rycr2P)

**MAJOR PROJECT 4:**

**ABC Call Volume Trend**

**Project Description:**

Dataset of a Customer Experience (CX) Inbound calling team for 23 days. 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).

**Inferences:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| •The Call volume follows a left skewed bell curve, with the 9588 at 9am to 10am peaking at 11 to 12. with 14626 then continuously declining to 5505 in 8pm to 9pm slot | | | | | | | | | | | | | | |
| •Overall decreasing trend is followed |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| •During the initial number of hours large number of calls are abandoned, and during the last hour . | | | | | | | | | | | | | | |
| •During the day more than 11 lakhs calls are received   |  | | --- | | •In morning hours from 9 am to 12 pm and from 6pm to 9pm the call duration is highest | | •Longest duration during 10am to 11am followed by 8pm to 9pm then 7pm to 8pm | | •lowest during 12pm to 1pm slot followed by 2pm to 3pm then 1pm to 2pm.  **LINK:** [**https://1drv.ms/x/s!AgVOadNMXixUoV7fGYhgVeD3vsxq?e=P9rEaB**](https://1drv.ms/x/s!AgVOadNMXixUoV7fGYhgVeD3vsxq?e=P9rEaB) | | | |  |  |  |  |  |  |  |  |  |  |  |  |

**CONCLUSION:**

In conclusion, I would like to tell that after doing a thorough analysis we were able to derive the insights from the data and was able to plot various graphs using that data. The data which once looked useless gave some very useful insights.

THE END