DATA ANALYTICS

PORTFOLIO PROJECT

PREPARED BY:

MOHAN CHANDRA

Contact:

Email ID: siliverumohanchandra@gmail.com

PROFESSIONAL BACKGROUND:

EDUCATION

I have done my Bachelors degree in Mechanical Engineering from JNTUH University Hyderabad in 2018 with an aggregate of 66.86%.

I worked as a Planning Engineer and BI Analyst in SMECH Engineering Services which is located in Maamba, the Southern Part of Zambia.

I have done Business Analyst using Power BI from EDUREKA, Advanced DAX for Microsoft Power BI Desktop in Udemy and SQL for Data Analytics in Intern Shala.

I'm Currently doing Business Intelligence Master's Program in EDUREKA.

SKILLS

- Structured Query Language (SQL)
- Microsoft Excel
- Data Visualization (Power BI and Tableau)
- Statistics
- DAX
- SSIS

ABOUT ME

Thorough and meticulous Data Analyst passionate about helping businesses succeed. Looking forward to work in an esteemed company where skills, knowledge, talents matter more than the degree. Through development of portfolio projects and internship, I learned the importance of having an iterative, hypothesis-oriented approach to analysis. I am eager to leverage that approach at your company as a data analyst.

DATA ANALYTICS PROCESS

Description:

We use Data Analytics in Everyday Life

6 STEP PROCESS

Plan:

I planned to watch movie and also to go to restaurant after watching the movie.

Prepare:

Next I need to prepare how much I'm willing to spend and to make proper arrangements.

Process:

Then I need to verify location and movie to watch based on IMDB Rating.

Analyze:

Next I want to go restaurant and to eat food based on the restaurant's famous dish especially vegetarian food.

Share:

Then I will ask restaurant owner to suggest their best vegetarian food.

Act:

Then I finally enjoy the day

Project Link:

https://docs.google.com/presentation/d/1eEEJNFrZkLGIkl_amdWG3YHGRj0fc8Rg/edit?usp=share_link&ouid=114843614286429142720&rtpof=true&sd=true

Instagram User Analytics

Project Description:

User Analytics is the Process by which we track how users engage and interact with our digital Product. The project is about finding out the various insights in Instagram User database and this helps to analyse the raw data to create useful insights. Various database management tools can be used to extract useful insights and even visualise them. This enables a way to increase efficiency of a platform and helping the business grow.

Approach:

I have analysed the database carefully. Observe all the tables, columns, rows, and relationship among all the tables. Then I have started to create tables in MySQL server. After finishing creating table, I have added the data into each table. Afterwards check all the tables' content carefully. SQL queries were used to create a database using the raw data provided.

Insights:

There are total of 100 users using Instagram clone.

Around 26% of the users are inactive in Instagram. We can remind the inactive users by sending them promotional emails to post their 1st photo.

The most liked photo in Instagram is posted by Zack_Kemmer93, which is liked by 48% of the users. The team can start the contest for the most liked photos. This will make the users to post more such good posts.

The most used hashtag is "smile". Around 59% of the users use the "smile" hashtags. If a partner brand use the "smile" hashtag, it will be able to reach the most users in the platform.

The best days to launch ads are Sunday and Thursday. As the most users register on Instagram on Sunday and Thursday.

13% of Instagram IDs are fake and dummy accounts.

Project Link:
https://drive.google.com/file/d/1Uus6UK5hFAXSuQ8CJHjeSQ3tyuA49pzo/view?usp=share_link

Operation Analytics and Investigating Metric Spike

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 and help them derive insights out of the data they collect.

Being one of the most important parts of a company, this kind of analysis is further used to predict the overall growth or decline of a company's fortune. It means better automation, better understanding between cross-functional teams, and more effective workflows.

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. Questions like these must be answered daily and for that its very important to investigate metric spike.

Iam working for a company like Microsoft designated as Data Analyst Lead and is provided with different data sets, tables from which you must derive certain insights out of it and answer the questions asked by different departments.

Approach:

I am working for a company like Microsoft designated as Data Analyst Lead and is provided with different data sets, tables from which I must derive certain insights out of it and answer the questions asked by different departments.

Firstly, I spent some time on understanding the data/table given. I cleared the questions which was in my mind and what are the things to consider while reviewing the data. I use SQL to derive different insights from the dataset provided by the management team. I first created a database "operation_analytics" and then the tables using the structure and links provided by the team. Then, we performed analysis to generate valuable insights for the company.

Insights: Case Study 1 – Job Data

- The number of distinct jobs reviewed per hour per day for November 2020 is 83%.
- We used the 7-day rolling average of throughput as it gives the average for all the days right from day 1 to day 7 whereas, daily metric gives the average for only that particular day itself.
- The percentage share of Persian language is the most (37.5%).
- There are two duplicate rows if we partition the data by job_id. But if we look the overall columns, all the rows are unique.

Case Study 2 – Investigating metric spike

- The weekly user engagement increased from week 18th to week 31st and then started declining from then onward. This means that some of the users do not find much quality in the product/service in the last of the weeks.
- There are in total 9381 active users from 1st week of 2013 to the 35th week of 2014.
- The overall count of weekly engagement per device used is the most for MacBook users and iPhone users.
- The email opening rate is around 34% and email clicking rate is around 15%. The users are engaging with the email service which is good for the company to expand.

Project Link:

https://drive.google.com/file/d/1ldVTQswKCpSgf5YYvHI94TBu1le6gO8o/view?usp=share_link

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 analyze before hiring freshers or any other individual. Thus, making an opportunity for a Data Analyst job here too!

• Being a Data Analyst, my job is to go through these trends and draw insights out of it for hiring department to work upon.

As an Analyst I required to use my knowledge in Statistics and use different formulas in Excel and draw necessary conclusions about the company.

Approach

• My Approach towards the Project mainly on understanding data columns and data, checking for missing data, checking for outliers and removing the outliers

• Excel plays the Vital role in Performing my entire analysis.

• Analyse data by using especially pivot charts and table.

• Solve the given problem step by step and create interactive graphs.

• Overall my execution works well for this project and done enough to obtain the useful insights for the company.

Insights

The rejection rate of male applicant is 6% higher than the female applicant.

The average salary paid in this company is 50K.

Most of the employers are in the Operation Department and then in the Human Resource Department.

The applicant is most likely to get hired if he/she is applying for the HR Department as the rejection rate here is the least.

There are only 3 candidates in the company who are paid more than 100K.

Project Link:

https://drive.google.com/file/d/13_-_5PHmJBSlqpugz61x3vcgtnWCAG7t/view?usp=share_link

IMDB Movie Analysis

Project Description:

The dataset provided by the company contains various columns of different IMDB Movies. We are required to Frame the problem. For this task, we will need to define a problem we want to shed some light on.

We can do this by asking 'What?'. This is where we frame the problem i.e. What is the problem?

We can do this by asking the following 'What?'

• What do we see happening?

• What is our hypothesis for the cause of the problem? (This will be broadly based on intuition initially)

• What is the impact of the problem on stakeholders?

• What is the impact of the problem not being solved? How to handle the things:

• Clean the data.

• Use the Data Analysis skills to explore the data set.

• Derive insights.

The things that we are going to find out through the project are movies with the highest profit, top movies as per IMDB rating, top directors, most popular genres, top foreign language films and more

Approach:

we used Five 'Whys' approach to determine its root cause by repeatedly asking the question “Why”. While asking Why is easy, what we're interested in is the answer. Each time we answer why the next time gets more difficult as we must think deeper behind the reasons for this. As we ask why, we may find that we have multiple answers for the same question.

This is the most important step to perform for the better analysis of the data.

- Dropped the unwanted columns as there is no use for the analysis.
- Dropped the rows which are having null/blank.
- Removed the duplicate row values.

Insights:

- There are as many as 5 outliers in the profit columns.
- The movie with the highest profit is ‘Avatar’
- The Shawshank Redemption is the top-most movie with the highest IMDB rating.
- The Good, the Bad and the Ugly (Italian) is the top-most foreign language movie.
- Charles Chaplin is the top-most director followed by Tony Kaye.
- The most popular genres is Drama followed by Comedy.
- 'Leonardo DiCaprio' is the critic-favourite as well as the audience-favourite actor.
- The most users voted in the decade 2000s and the least in the decade 1940s.

Project Link

https://drive.google.com/file/d/1XDlEM48UHqX7Wt3OjkcMRs0rx_FI5AW/view?usp=share_link

Bank Loan Case Study

Project Description:

This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

Business Understanding:

The loan providing companies find it hard to give loans to the people due to their insufficient or nonexistent 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:

Identify the missing data and use appropriate method to deal with it. (Remove columns/or replace it with an appropriate value) Given that the file contained a large number of columns, it is important to clean and eliminate unnecessary data before analyzing it.

The steps involved in cleaning the data include.

- Eliminate columns that do not contribute to the business scenario.
- Remove columns with null values >30%.
- Impute rows with some value for columns with a low percentage of null values.
- Removing outliers from the data set.

Data Imbalance:

- To identify the imbalance, we must create a client count plot with targets of 1 and 0.
- Target variable (0 - all other situations, which we can name non-default scenarios for ease of use), 1 - client with payment difficulties: he/she had la ate payment of more than X days on at least one of the first Y installments of the loan in our sample.
- The proportions of 0 and 1 are, respectively, 91.92 and 8.07 This data collection is quite unbalanced.

• Analysis: We will produce several data sets with target values of 0 and 1 and then determine the correlation between various variables for both data frames with target values of 1 and 0.

"From our observations, we can see that individuals who consistently pay back their loans on time are offered higher credit limits by the bank, as opposed to those who have a history of payment issues."

RESULT:

From the analysis of the data, we can see that the majority of clients opt to repay their loans using the 'Cash through the bank' option, while the 'non-Cash from your account' & 'Cashless from the account of the employee options are not widely used. Additionally, most loan applications are for 'Cash loans and 'Consumer loans', but cash loans have a higher rate of being rejected.

A large portion of loan applications come from repeat customers, with 70% being from repeat clients, however, these customers are also more likely to have their loans rejected.

The loan applications for individuals with lower AMT_ANNUITY are more likely to be canceled or unused, and similarly, loan applications with a high AMT ANNUITY also have a higher rate of being rejected. Additionally, loan applications with a low credit amount are also more likely to be canceled or unused.

The clients who have been approved for a loan in the past are less likely to have issues with loan repayment as compared to those who were previously denied.

Project Link:

https://drive.google.com/file/d/1lQPkAOqed5ti8JwwOz_uBpfvMulqZGyc/view?usp=share_link

XYZ Ads Airing Report Analysis

Project Description:

Advertising is a way of marketing your business in order to increase sales or make your audience aware of your products or services. Until a customer deals with you directly and actually buys your products or services, your advertising may help to form their first impressions of your business. Target audience for businesses could be local, regional, national or international or a mixture. So they use different ways for advertisement. Some of the types of advertisement are: Internet/online directories, Trade and technical press, Radio, Cinema, Outdoor advertising, National papers, magazines and TV. Advertising business is very competitive as a lot of players bid a lot of money in a single segment of business to target the same audience. Here comes the analytical skills of the company to target those audiences from those types of media platforms where they convert them to their customers at a low cost.

Dataset having different TV Airing Brands, their product, their category. Dataset includes the network through which Ads are airing, types of network like Cable/ Broadcast and the show name also on which Ads got aired. You can also see the data of Dayparts, Time zone and the time & date at which Ads got aired. IT also includes other data like Pod Position (the lesser the valuable), duration for which Ads aired on screen, Equivalent sales &, total amount spent on the Ads aired.

Findings:

What is Pod Position? Does the Pod position number affect the amount spent on Ads for a specific period of time by a company? What is the share of various brands in TV airings and how has it changed from Q1 to Q4 in 2021?

Conduct a competitive analysis for the brands and define advertisement strategy of different brands and how it differs across the brands.

Mahindra and Mahindra wants to run a digital ad campaign to complement its existing TV ads in Q1 of 2022. Based on the data from 2021, suggest a media plan to the CMO of Mahindra and Mahindra. Which audience should they target?

Approach:

- Scatter chart w.r.t different brands is used to know if the pod position affects the amount spent on Ads for a specific period of time by the company.
- Bar chart and column chart are used for answering the share of various brands in TV Airings.
- We used pivot table to conduct the competitive analysis for the brands.

- Clustered Column chart is used to suggest a media plan to the CMO of Mahindra and Mahindra

Insights:

- The money spent by Mahindra and Mahindra is the most for the pod position ads (40.1%) and the money spent by Honda Cars is the least for the pod position ads (3.5%).
- The money spent by the Maruti Suzuki is the most consistent for all the Quarters of the year.
- People watch the most in the prime time and on weekend.
- The brand's money spent for the advertisement is the least for the last quarter pod position and the highest for the first quarter pod position.
- The Ads are shown the least in the prime access and evening news parts of the day.

Project Link:

https://drive.google.com/file/d/16wh4OHlmBgq6ZbC2o0I20HFgD6Q6L8_S/view?usp=share_link

ABC Call Volume Trend Analysis

Project Description:

A customer experience (CX) team consists of professionals who analyze customer feedback and data, and share insights with the rest of the organization. Typically, these teams fulfil various roles and responsibilities such as: Customer experience programs (CX programs), Digital customer experience, Design and processes, Internal communications, Voice of the customer (VoC), User experiences, Customer experience management, Journey mapping, Nurturing customer interactions, Customer success, Customer support, Handling customer data, Learning about the customer journey.

Let's look at some of the most impactful AI-empowered customer experience tools you can use today:

Interactive Voice Response (IVR), Robotic Process Automation (RPA), Predictive Analytics, Intelligent Routing

In a Customer Experience team there is a huge employment opportunities for Customer service representatives A.k.a. call centre agents, customer service agents. Some of the roles for them include: Email support, Inbound support, Outbound support, social media support.

Inbound customer support is defined as the call centre which is responsible for handling inbound calls of customers. Inbound calls are the incoming voice calls of the existing customers or prospective customers for your business which are attended by customer care representatives. Inbound customer service is the methodology of attracting, engaging, and delighting your customers to turn them into your business' loyal advocates. By solving your customers' problems and helping them achieve success using your product or service, you can delight your customers and turn them into a growth engine for your business.

Findings:

- The average call time duration for all incoming calls received by agents (in each Time_Bucket).
- The total volume/ number of calls coming in via charts/ graphs [Number of calls v/s Time].
- 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 in a day[9 pm to 9 am]. Maximum Abandon rate assumption would be same 10%.

Assumption: An agent work for 6 days a week; On an average total unplanned leaves per agent is 4 days a month; An agent total working hrs is 9 Hrs out of which 1.5 Hrs goes into lunch and snacks in the office. On average an agent occupied for 60% of his total actual working Hrs (i.e 60% of 7.5 Hrs) on call with customers/ users. Total days in a month is 30 days.

Insights:

- ✓ The company can hire 17 customer support agents for the night shift work.
- ✓ The customers call the least in the evening. So, the company can reduce the number of agents at that time for answering the calls.
- ✓ The company can shift some of the day workers for the night shift.
- ✓ The employees who are working 9 am to 9 pm. The manager can change some of the workers shift from 5 am to 2 pm and some workers from 2 pm to 11 pm to get the most calls answered.
- ✓ The company can make the employers divide into 3 parts too, so that the agents are always available 24/7.

Project Link:

https://drive.google.com/file/d/1CdHxqaGYb6XvBnz1zDpLQtdNv0tU39Y/view?usp=share_link

CONCLUSION:

Finally, these projects enhance my skills effectively. The things I learned from these projects are Predictive Analytics, HR Analytics, Risk Analytics, Behavioural Analytics, Customer experience and inbound Customer Support, analyse the huge datasets using excel, pivot table etc.,, SQL and Advance SQL Concepts, visualise the data to get valuable insights, concepts of operation analytics and investigating metric spike and many more.

However, these projects are almost related to real time business scenarios. So, this will definitely help me in future aspects. These Projects helps me analyse how company takes quick decisions by using raw data to convert in to valuable insights