Data Analysis Portfolio

PREPARED BY : -

AJAY LAD

Professional Background

I am pursuing my graduation in MSC(Computer Science). well currently I'm pursing Data Analyst course through Udemy and learned several skills like Python , Excel, SQL, Data Analysis , Power BI and Machine Learning.
Currently I am working as an intern at trainity and also pursuing my Data analyst where I am handling big data which include data analysis, data cleaning and identifying key insight into business in way through data can be used to solve problem and help in the growth of company.
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

IMDB Movie Analysis

We are providing you with dataset having various columns of different IMDB Movies. You are required to Frame the problem. Use these questions to guide your thinking: What do you see happening?

- What is your 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?

Problem

Once you have defined a problem, clean the data as necessary, and use your Data Analysis skills to explore the data set and derive insights.

Make sure that you have framed the problem and gathered initial insights from the data, you can ask the following questions as you dig deeper into your analysis.

Once you have the problem better defined, you can use 5 Whys technique to determine its root cause by repeatedly asking the question that "WHY" is this happening.

It's also called the Root Cause Analysis, developed by Sakichi Toyoda, founder of Toyota Industries. Here's an example of how this technique could be used to figure out the cause of the following problem: A business went over budget on a recent project.

While asking Why is easy, what we're interested in is the answer. Each time you answer why the next time gets more difficult as you must think deeper behind the reasons for this. As you ask why, you may find that you have multiple answers for the same question.

Design

Steps taken to clean the data:

- First importing the datasets provided and merging them.
- Then removing the duplicates and the blank cells.
- Improving the headers of each column with proper values.

Use this question for guidance:

- What do you see happening?
- What is the impact of the problem on stakeholders?
- What is the impact of the problem not being solved?

Finding-1

A. **Movies with highest profit : Create a new column called profit which contains the difference of the two columns: gross and budget. Sort the column using the profit column as reference. Plot profit (y-axis) vs budget (x- axis) and observe the outliers using the appropriate chart type.**

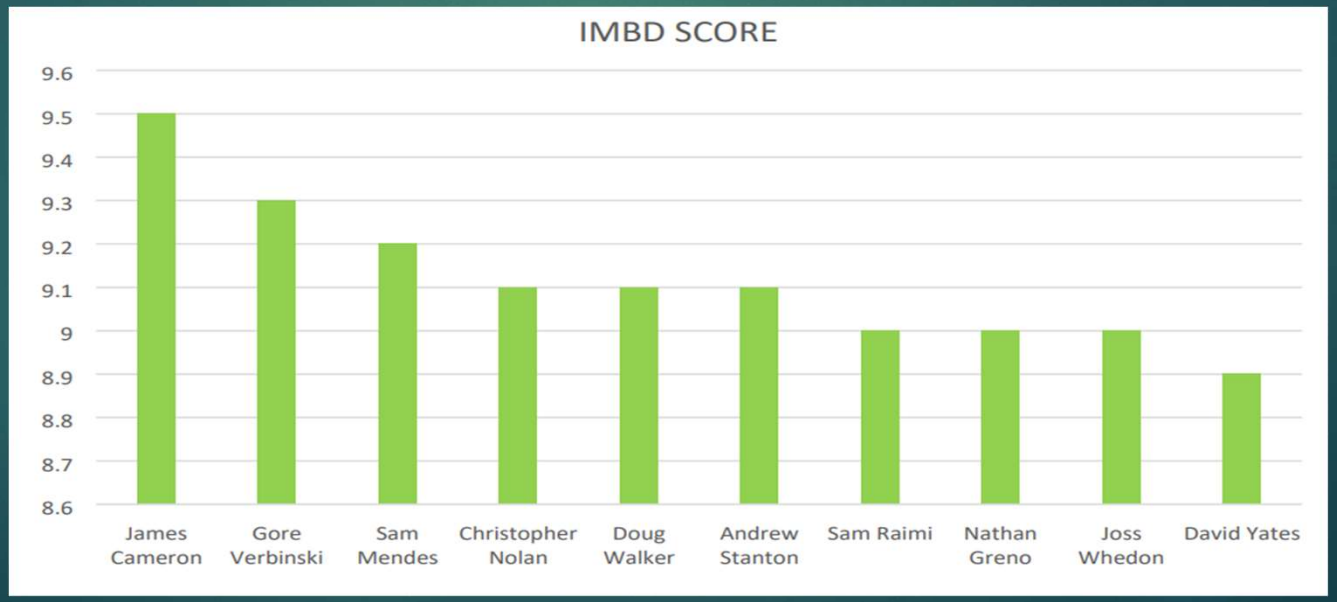
Your task : Find the movies with the highest profit?

Ans. Gross – Budget = Profit Highest earning movie is Avatar

Gross	Budget	Profit
7.16E + 08	2.37E+08	523505847

Finding-2

B. Best Directors: TGroup the column using the director_name column. Find out the top 10 directors for whom the mean of imdb_score is the highest and store them in a new column top10director. In case of a tie in IMDb score between two directors, sort them alphabetically.
Your task: Find the best directors



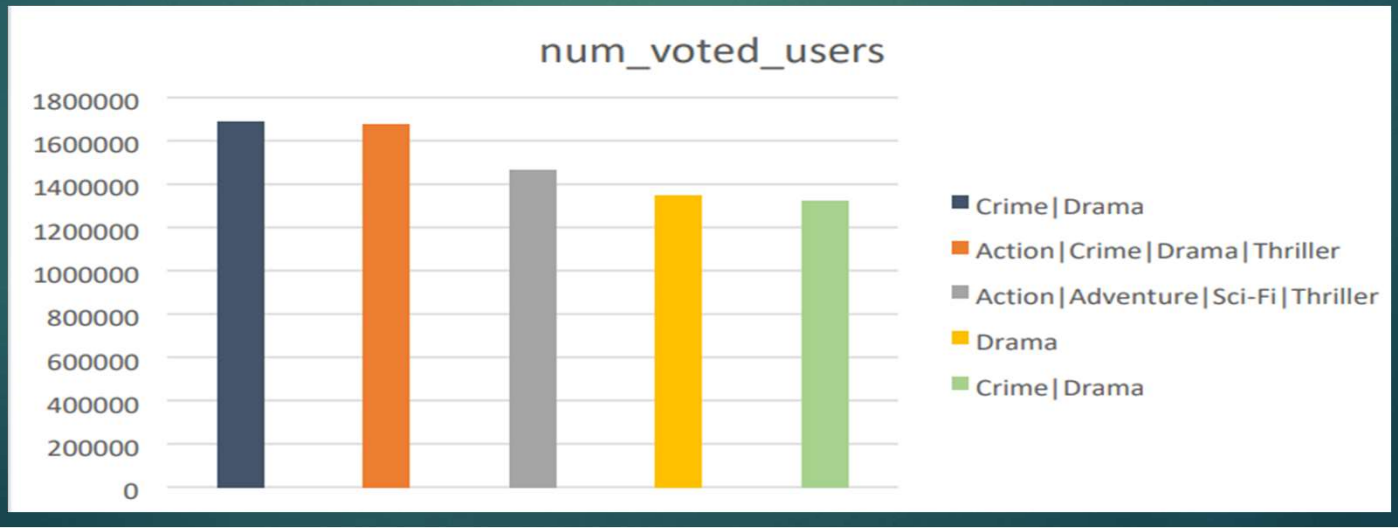
Best director on the base of imdb score : James Cameron with 9.5 imdb score .

Finding-3

C. Popular Genres : Perform this step using the knowledge gained while performing previous steps.

Your task : Find popular genres

Ans. Top 5 Genres



So, while analyzing the data we can clearly opt out that Crime Drama is the most voted Genres with 1689764 votes.

Analysis

After the analysis, I found out that "best director" on the base of imdb score is James Cameron among all the directors.

Five 'Whys' approach help me to find the root cause: -

Q: Why did we go over budget on our project?
A: It took much longer than we expected to complete.

Q: Why did it take longer than expected to complete?
A: We had to redesign several elements of the product.

Q: Why did we have to redesign elements of the product?
A: Features of the product were confusing to use.

Q: Why were the features of the product confusing to use?
A: We made incorrect assumptions about what users wanted.

Q: Why did we make incorrect assumptions about what users wanted?
A: Our user experience research team didn't ask effective questions.

Conclusion

To summarize, based on my analysis, movies with drama, action, and crime genres are the most popular. In conclusion, the highest gross for movie is Avatar with outstanding profit of $ 52.35 Cr which I had plot various graphs using that data.

To fine-tune our strategy, further research will need to be done to determine which movies to stream, which ones to create, and which director is directing it. If we are strategizing for international growth, then we will need to analyse viewer's interests in different emerging markets as well.

Call Volume Trend Analysis

We are providing you with a dataset of a Customer Experience (CX) Inbound calling team for 23 days. Data includes Agent_Name, Agent_ID, Queue_Time Time, Time_, Duration, Call_, call status (Abandon, answered, transferred).

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, 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.

Design

First and most important step is to clean the raw data :

- First importing the dataset provided.
- Then removing the duplicates and the blank cells.
- Improving the headers of each column with proper values.

Tools used for visualization :

- Excel

Finding-1

A. Calculate the average call time duration for all incoming calls received by agents (in each Time Bucket).

Ans. (Total call received (in sec) – IVR duration) / Total call received (in sec)

Deducting IVR duration because that is automatic call answer by the AI and that is not attended by the agent.

= (16463119.00 – 58580) / 83791
= 195.77 sec
= 3.26 min is the avg time call

Finding-2

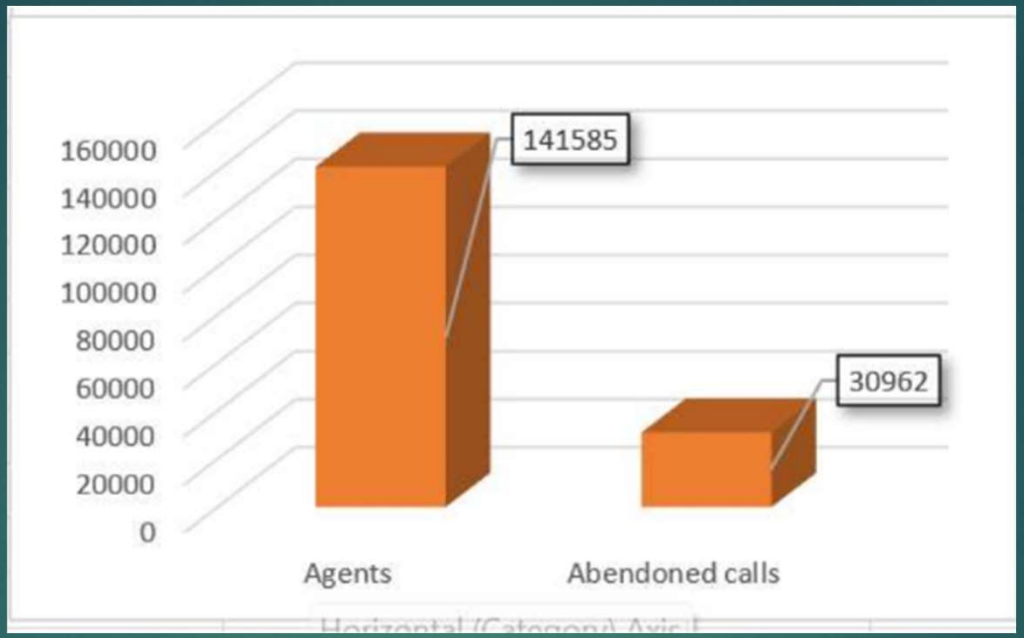
B. As you can see current abandon rate is approximately 30%. Propose a manpower plan required during each time bucket [between 9am to 9pm] to reduce the abandon rate to 10%. (i.e. You have to calculate minimum number of agents required in each time bucket so that at least 90 calls should be answered out of 100.)

Ans. Total abandon calls = 34,404
Total number of calls = 1,17,976

% Of abandon calls = (Total abandon calls / Total number of calls) * 100
= (34404/11797) *100 = 29.161 %

Approx. 30% of abandon ratio

- To reduce the abandon calls ratio to 10% we can increase the man power to 20% so that maximum number of calls can be attend.
- Total number of executives are 83791
- So here increasing it to 20% we will get 16,758 17
- Total number of executives after increasing 20% = 1,00,549



- Series 1 indicate executives working as per data.
- Series 2 indicate executives working after purpose of manpower.

Finding-3

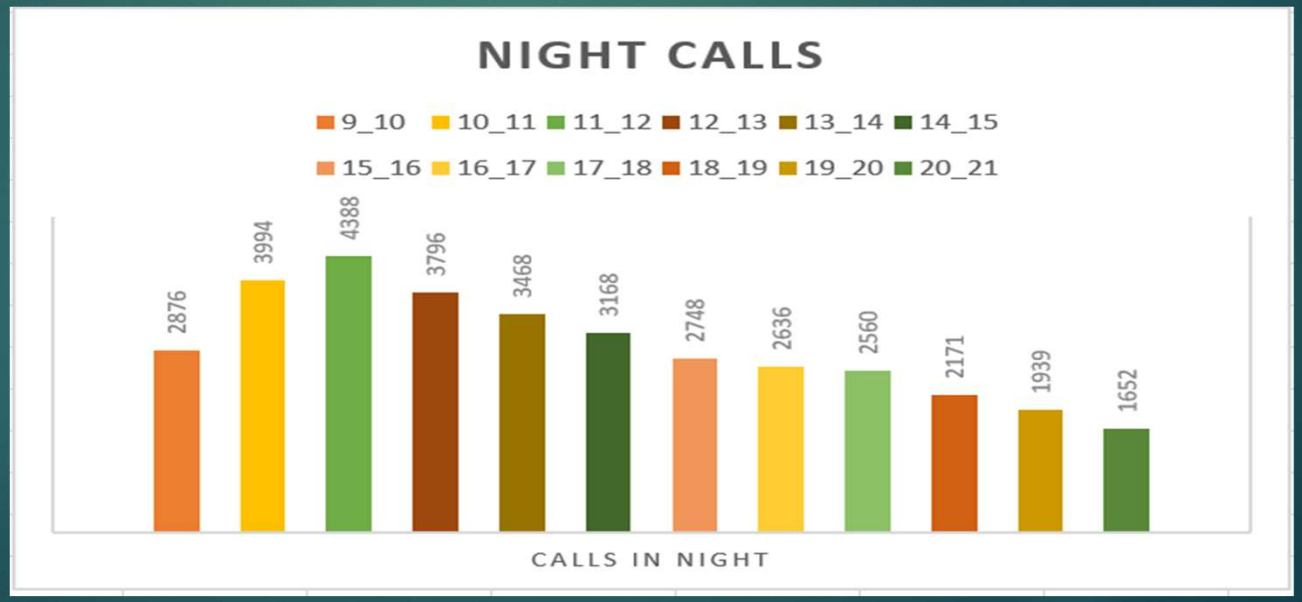
D. Let's say customers also call this ABC insurance company in night but didn't get answer as there are no agents to answer, this creates a bad customer experience for this Insurance company. Suppose every 100 calls that customer made during 9 Am to 9 Pm, customer also made 30 calls in night between interval [9 Pm to 9 Am] and distribution of those 30 calls are as follows:

Time_Bucket	9_10	10_11	11_12	12_13	13_14	14_15	15_16	16_17	17_18	18_19	19_20	20_21
Calls In Night	2876	3994	4388	3796	3468	3168	2748	2636	2560	2171	1939	1652

Now propose a manpower plan required during each time bucket in a day. Maximum Abandon rate assumption would be same 10%.

Ans. From the question suppose every 100 calls that customer made during 9 Am to 9 Pm, customer also made 30 calls in night between interval [9 Pm to 9 Am] 19

- From the graph we can say that highest call received is 5 in between 8 to 9 am, so from analyzing the data we can say that we have to hire executive for night shift we can be maximum to 5- 6.



Second suggestion can be like if we don't want to spend extra for night executive then we can shift some of day employee shift to night as it only required only 5-6 maximum executive.

- Or as said in question that company can consider 10% abandon So, with that data we can add 3-4 executive manpower.

Analysis

After the analysis, I got know to that company is struggling with manpower, moreover current Abandon rate of company is near around 30% which they want to reduce it to 10%.

Five 'Whys' approach help me to find the root cause: -

Q. Why Abandon ratio is high?
A. This causing because of less of manpower.

Q. Why is there is less of manpower?
A. Because of low manpower.

Q. What Business challenges does the company face because of low manpower?

A. Due to low manpower abandon ratio is increasing and company isn't able to provide service to its customer.

Q. What are the steps company have to take to reduce abandon ratio?

A. Company have to hire more employee for morning as well as night shift because from the data it is visible that customer is also calling between 9pm – 9am.

Q. What % of employee do company need to increase to reduce abandon rate to 10%?

A. Near about 20% of current employee they have to increase to reduce the rate to 10%.

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.

Appendix

Google Drive Link for All Projects:-

https://drive.google.com/drive/folders/1c6AJl72lGJ28hbMU9ToeR9652tWSGOEh?usp=share_link