PORTFOLIO PROJECT

By – Samiksha Zagade

Abstract

Summary of all the projects I have completed.

Samiksha Zagade

Samizagade30.gmail.com

ABOUT ME

I am Samiksha Zagade. I have a Bachelors Degree in BSc Mathematics and currently I'm pursuing a PG Diploma Course in Data Science from IIIT Bangalore. I'm a fresher and I have great interest in data science and data analytics which lead to me enrolling for the course offered by trainity in Data Analytics which has helped me in building my portfolio and skills too.

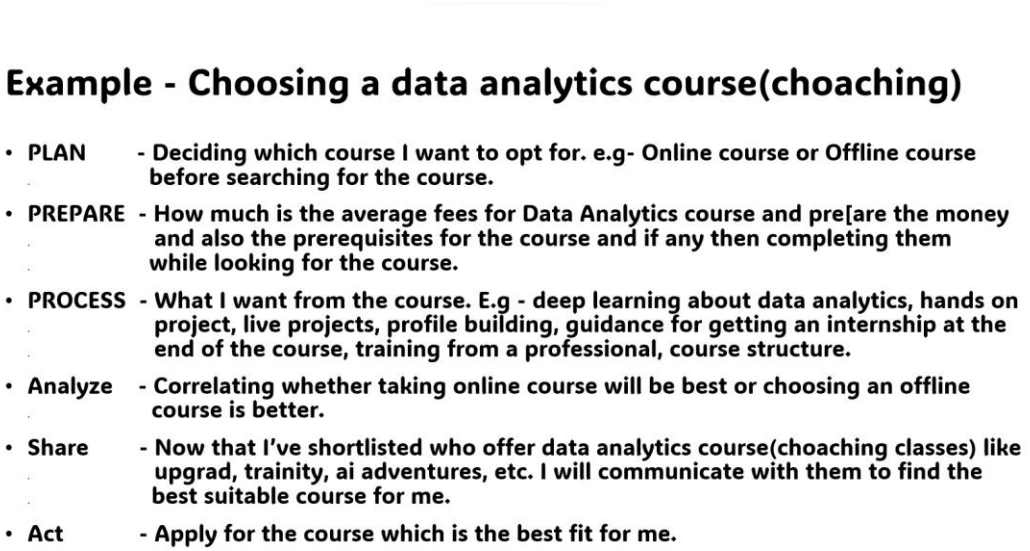
I have listed below all the projects that I have worked on, its description and the insights that I have gained while working on those respective projects. Also the tools used are mentioned along with the project information.

TABLE OF CONTENTS

Module 1: Data Analytics Process

Overview

Explored the entire data analytics process, from data collection to insights generation.



Key Achievements

- Proficiency in data cleaning techniques.
- Mastery of data visualization for effective communication.
- Fundamental skills in exploratory data analysis.

Module 2: Instagram User Analytics

Overview

Conducted a comprehensive analysis of user engagement and content performance on Instagram. This project aims analyzing user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow. In this project, I have used SQL and MySQL Workbench as a tool to analyze Instagram user data and answer questions posed by the management team. I have tried getting the answers to the questions raised. The insights derived from this analysis are used by various teams within the business.

Tools Used – Excel, Power Point

INSIGHTS

- After analyzing this data we were able to answer important questions which lead to valuable insights.
- Even though there were many loyal users but we found considerable amount of users were inactive about 26 users were inactive.
- Andre_Purdy85 username had the most likes and hence won the contest.
- Five most hashtags used were sunrise, sunset, stunning, smile and style. Thus these hashtags will enable the post to reach maximum number of people as these are the most popular hashtags.
- Thursday and Sunday are the days when most users register.
- The average number of posts per user is 3.4730. The average tells if users are still active or not. Thus many users are still very active and their response is very good.
- There are in total 13 such accounts registered which are fake or dummy accounts .

RESULTS

- Thus the list of these inactive users will be encouraged to use Instagram by sending them emails.
- Holding contest like this will not only motivate the winners to be more active but also increase the competitiveness of the users which will increase their activity on Instagram. Thus holding such contest will attract more potential users and inactive users too.
- Using hashtags helps reach out to large crowd with ease and this helps with marketing and maximizes the view of the post.
- Instagram team can launch new ads on these days. Every Thursday and Sunday new ads and contest can be posted which will attract many users.
- The average number of posts per user tells if users are still active or not. Thus many users are still very active and their response is very good.
- The fake accounts can either be restricted or blocked so that they do not cause inconvenience to other users.
- Making the policies strict can also help in decreasing the number of fake or dummy accounts. These accounts can damage the application rating as the users will be inconvenienced by these accounts which something we want to avoid.
- The project has helped in finding the Instagram team where they were lacking and how some changes they had made affected the progress which has been good for example holding contest, and engaging the inactive users.

Key Achievements

- In-depth understanding of social media analytics.
- Identification of key engagement metrics.
- Visualization of user behavior patterns.

Module 3: Operation & Metric Analytics

Overview

- Explored operational metrics to optimize business processes and enhance efficiency. In this project our key aspect of Operational Analytics is investigating metric spikes. 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. In this project, I have used SQL and MySQL Workbench as a tool to derive insights from this data to answer questions posed by different departments within the company. We have segmented the analysis in two parts – Case study 1 and Case Study 2. The questions posed by the team and the insights are as follows for the respective Case studies.

Insights and Results

- Working on the Case study 1 we gained some insights on the data.
- Maximum jobs were reviewed on 28 November 2020.
- 7-day rolling average for throughput is prefered over daily metric because 7-day rolling average smooths out the fluctuations in the data and provides a more accurate representation of the overall trend.
- Persian Language has highest percentage (37.5%) in the last 30 days in comparison to other languages having 12.5%.
- The actor id 1003 had two rows and was the only duplicate column.

Insights and Results

- Next we worked on the Case study 2.
- Friday had the highest involvement when we found the weekly user engagement.
- December has the highest active user in 2013 but in 2014 August had the highest active users.
- The highest user activeness was the device device macbook pro.
- Sent_weekly_digest was the user mostly engaging with the email service.

Tools Used – Excel, Power Point

Key Achievements

- Utilized metrics to assess and improve operational workflows.
- Applied data-driven decision-making in process optimization.
- Developed skills in metric analysis for business operations.

Module 4: Hiring Process Analytics

Overview

- Analyzed hiring data to enhance the efficiency and effectiveness of the recruitment process. The objective of this project 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. I have been given a dataset containing records of previous hires. After analysing the data I have answered certain questions that can help the company improve its hiring process. I have used Microsoft Excel 2019 to do this project.

Insights and Results

- Majority of males get hired compared to females which we visualised in the bar chart also.
- The average salary of the company is Rs.49873 approx.
- The highest number of people lie in the salary range 40000 to 60000.
- Operations Department has the highest number of people working in that department and Hr department has the least number of people working in that segment.

Tools Used – Excel, Power Point

Key Achievements

- Evaluation of candidate sourcing channels.
- Identification of bottlenecks in the hiring pipeline.
- Application of analytics to improve the quality of hires.

Module 5: IMDB Movie Analysis

Overview

Conducted a thorough analysis of movie data from IMDB, exploring trends and insights. The dataset 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 or IMDB score that is assigned to each movie. 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.

Insights

- Comedy, Action and Drama are in more demand and common genres.
- Directors can prefer making a movie of Comedy genre as this genre has highest demand by the audience.
- The duration between 75 and 150 has maximum count of ratings ranging from 8 to 2 approx. Highest variation is seen between this duration range.
- The highest viewed movies is of English language. The movies with English language will have a big audience which will increase the reach of the movie to maximum people.
- So making a movie of English language will be very beneficial.
- The directors like Akira Kurosawa, Alfred Hitchcock and Asghar Farhadi have the making average rating for their movies. Thus working with these top three directors has high percent of movie success.
- If these directors make a movie in English then maximum profit will be gained.

Tools Used – Excel, Power Point

Key Achievements

- Proficient in data extraction from diverse datasets.
- Visualization of trends in movie preferences.
- Application of statistical methods for trend analysis.
-

Module 6: Bank Loan Case Study

Overview

Analyzed factors influencing bank loan approvals and assessed credit risk. The main aim of this project is to identify patterns that indicate if a customer will have difficulty paying their instalments. This information can be used to make decisions such as denying the loan, reducing the amount of 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.

Results

- Common area average, common area mode and nonliving apartments median have around 89% missing values which means more than half of the data is missing.
- AMT_CREDIT, AMT_ANNUITY, AND CNT_FAM_MEMEBERS are observed to have outliers. The outliers were dealt with as outliers can significantly impact the analysis and distort the result
- NAME_CONTRACT_TYPE, CODE_GENDER and TARGET are the columns which have data imbalance.
- Many customers came accompanied when applying for the loan.
- Most loyal customer who apply for the loan are married and they have high count of repaying the loan in time.

- Hence the married customers can be given loan benefits for being loyal customers.
- Very few students and unemployed people apply for loan, to attract more of these customers the bank should make new policies which will benefit them as well and enable the bank in getting more customers.
- Most of the customers of the bank have income between the range 25000-275000.
- Females having secondary / secondary special education apply for loans more.
- Sunday has very less customers having payment difficulties.
- Tuesday has highest count of 0 as target variable indicates that the payment was made on time followed by Friday.
- Highest number of customers are labourers who pay the loan back without difficulties and on time.
- Highest number of customers are labourers who pay the loan back without difficulties and on time.
- GOODS_PRICE and AMT_CREDIT have very high positive correlation the value being 0.987 and DAYS_EMPLOYED and REGION have very low negative correlation.

Tools Used – Excel, Ms Word

Key Achievements

- Evaluated credit risk using financial analytics.
- Applied predictive modeling to assess loan approval likelihood.
- Enhanced skills in financial data analysis.

Module 7: Impact of Cars Features

Overview

Evaluated the effectiveness of advertisements by analyzing car features. The main aim of this project is to analyse 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.

Tools Used – Excel, Ms Word

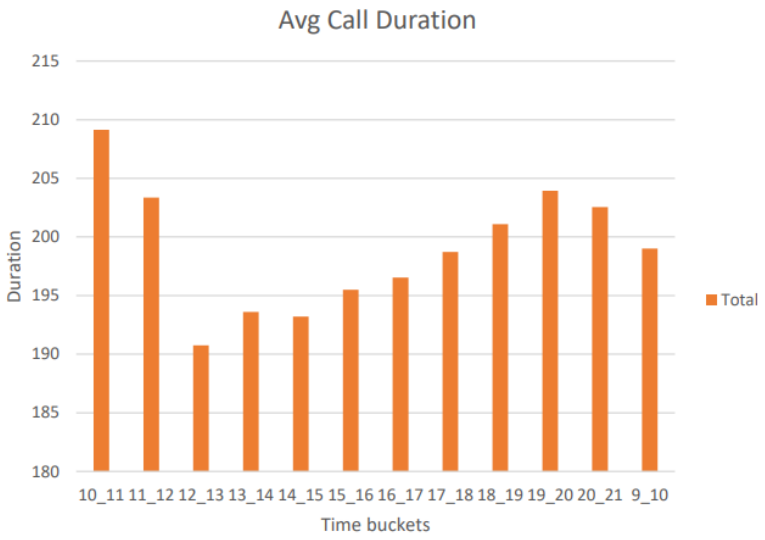
Key Achievements

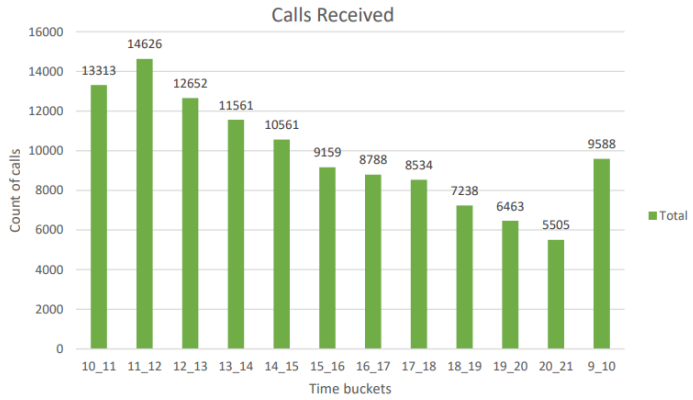
- Assessment of campaign performance metrics.
- Utilized analytics to optimize advertising strategies.
- Visualization of ad performance for stakeholders.

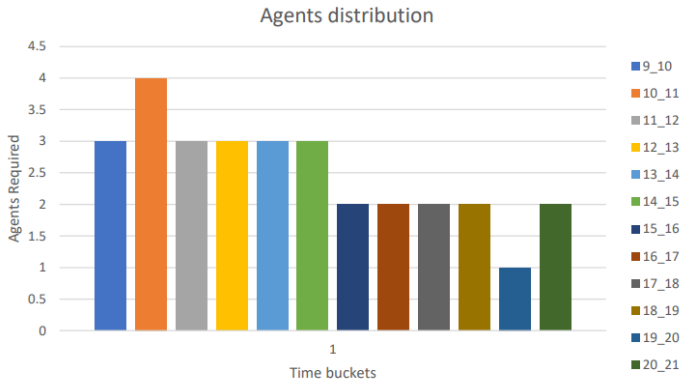
Module 8: ABC Call Volume Trend Conclusion

Overview

Investigated call volume trends to optimize customer service operations. This project mainly focuses on Customer Experience (CX) analytics, specifically focusing on the inbound calling team of a company. The dataset spans over 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). The aim is to analyze customer feedback and data, derive insights from it, and share these insights with the rest of the organization. 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. The goal is to attract, engage, and delight customers, turning them into loyal advocates for the business.







Tools Used – Excel, Power Point

Key Achievements

- Analyzed call volume patterns for service optimization.
- Utilized analytics for resource allocation in customer service.
- Enhanced skills in customer service analytics.

Overall Conclusion

- **Skill Development:** Developed a diverse skill set, encompassing data cleaning, visualization, and advanced analytics techniques.
- **Real-world Application:** Applied analytics to real-world scenarios in various domains.
- **Problem-solving Proficiency:** Gained expertise in deriving actionable insights to solve complex business problems.