Google capstone project report

Delight Ogwor

Please note that although this project was done using, power query for cleaning and organizing and power bi for analysis and visualization.

Case Study: How Does a Bike-Share Navigate Speedy Success?

The purpose of this document is to consolidate downloaded Divvy data into a single data frame and then conduct a simple analysis to help answer the key question: "In what ways do members and casual riders use Divvy bikes differently?"

Introduction

This exploratory analysis case study is towards the Capstone project requirement for Google Data Analytics Professional Certificate. The case study involves a bike-share company's data on its customer's trip details over 12 months (January 2022 - December 2022). The data has been made available by Motivate International Inc. under this license. The analysis will follow the 6 phases of the Data Analysis process: Ask, Prepare, Process, Analyze, and Act. A brief explanation of these processes:

Act

• Ask effective questions

• Define the scope of the analysis

• Define what success looks like

Prepare

• Verify data's integrity

• Check data credibility and reliability

• Check data types

• Merge datasets

Process

• Clean, Remove, and Transform data

• Document cleaning processes and results

Analyze

• Identify patterns

• Draw conclusions

• Make predictions

Share

• Create effective visuals

 • Create a story for data

• Share insights with stakeholders

Act

• Give recommendations based on insights

• Solve problems

• Create something new

The scenario: The marketing team needs to design marketing strategies aimed at converting casual riders into annual members. To do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ.

Stakeholders:

 • Director of Marketing

• Cyclitic executive team

Objective: Hence, the objective of this analysis is to throw some light on how the two types of customers: annual members and casual riders, use Cyclistic bike-share differently, based on a few parameters that can be calculated/ obtained from existing data.

Deliverables: • Insights on how annual members and casual riders use Cyclistic bikes differently

 • Provide effective visuals and relevant data to support insights

• Use insights to give three recommendations to convert casual riders to member riders

Cleaning process

I cleaned the data set with power query in power bi after I discovered the dataset was too large for excel. I loaded the folder containing the 12 datasets on the power query

I started by deleting the ride id, longitude, and latitude as I figured they won't be necessary.

Since the date column contained time, I extracted it so that I would be able to know the hours the trip started and the duration of each ride,

A significant part of the data for June was blank and since there was no way to ask if this can be found, I proceeded to drop the rows that contained blank and null values and renamed the columns for better understanding.

For visualization and findings, pls check the other pdf file.