Roller Coaster Capstone Project

Tech Stack

Tableau, Python, Pandas, Jupyter Notebook, Selenium, SQL

Introduction

Welcome to your biggest challenge of the course! This capstone is a case study for a simulated amusement park company that operates globally called **Family Amusement, Inc.**

Each year, the *Amusement Today* publication does the **Golden Ticket Awards**, which bestows awards in a variety of categories to the top amusement parks in the world and their roller coasters. Family Amusement has yet to break into any of the categories and is hopeful that 2023 will be the year they can achieve one of the coveted awards.

Family Amusement is planning to start construction on their most ambitious amusement park yet, but before they break ground, they need to make sure they are making the wisest decisions on a few key points in the process.

---

1. ***Where*** *should the park be built?*
2. *What* ***metrics*** *should guide the construction of a* ***flagship roller coaster*** *that can compete in ranking with other highly ranked roller coasters in the same category?*
3. *Which* ***material(s)*** *should be used to build the coaster? (Wood, steel, or a hybrid of both).*

Using data that we will supply, data that you will scrape or otherwise obtain, and all of the analytics tools in your toolbox, you will create a compelling case for your recommended answers to the above questions, as well as coming up with a few questions of your own that are worth consideration in the process.

Key Points

- **The training wheels are off!** Instructors will be here to help guide you as you move through the process of performing this data analyzation, but what conclusions you come to and how you get there will be up to you.
- We've supplied a slightly cleaned data set that will need to be cleaned depending on how you need to use it. ***Before moving data into Tableau, we recommend completely cleaning in Pandas/Jupyter Notebook and merging datasets there.***
- A website has been provided to the **Golden Ticket Awards** that has useful additional information. Some of that information may even be data you decide to scrape!
- The process of coming to your conclusions is not often linear. A typical analysis project will start with data collection, move to data cleaning/wrangling, then go into a visualization/reporting step. Analyzation happens during the entire process, because you're always looking for possible patterns that lead to insights you can use to answer the business questions. **Be sure to document as you go along!**

Capstone Materials

MySQL Database Connection String

- **host:** **"devcodecamp-mysql.c5p7yvjq1h8y.us-east-2.rds.amazonaws.com"**
- **database:** **rollercoasterdb**
- **tables:**
    - **coaster_metrics**
    - **park_locations**
- **user:** **"root"**
- **password:** **"D3v$tudent"**
- **use_pure:** **"True"**

Golden Ticket Winners Website

https://goldenticketawards.com/past-gta-winners/

Starter Materials Repository Link

https://github.com/dCC-Online/Capstone_DA_Starter

User Stories

Total Weighted Project Points: /245

Main Stories

- **(25 points):** As a data analyst, I want to collect any data that will be useful for proving or disproving my proposed course of action using any means available (curating existing data sources in SQL or CSV/Excel format, pulling data from APIs, scraping data from relevant websites where legal to do so).
- **(25 points):** As a data analyst, I want to clean my data sources and wrangle them into a useful form for analysis. This includes *removing null values, duplicates, outliers, reformatting column names, and creating calculated columns as needed.*
- **(50 points):** As a data analyst, I want to build a Tableau workbook that includes a **MINIMUM** of the following:
    - 4 Worksheets (4 individual visualizations)
    - 2 Dashboards
        - At least one Dashboard should contain at least 1 Action
    - 1 Story with at least 5 Story Points
        - Be sure to state the **business questions** and choose to either present your conclusion at the beginning or end of the Story.

- **(25 points):** As a data analyst, I want to define and clearly state a hypothesis and proposed course of action that would **answer the case study questions** stated in the Introduction.
- **(45 points):** As a data analyst, I want to pose **ONE** additional evaluation question based on research and analysis of the data, then incorporate those questions and their answers into my Story.
- **(75 points):** As a data analyst, I want to **present my completed Tableau Story** to a live instructor as part of a **Level III reporting structure (formal synchronous)** that will include a case made for your chosen conclusions, proposed courses of action (if any), and additional training on how to use the completed Tableau workbook in case the stakeholder(s) present would like to drill down further on your analysis.

Checklist

🔵 ***Analysis happens at all stages! You are actively seeking connections and generating insights as you go.*** 🔵

- ☐ Inspect, clean, and wrangle the existing data from SQL and CSV files.
- ☐ Initial analysis in Pandas and Tableau.
- ☐ Research answers to evaluation questions and pull together any supporting references.
- ☐ Scrape any additional data you deem useful and research/utilize possible APIs.
- ☐ Fold new data into a data set and finalize visualizations, dashboards in Tableau.
- ☐ Construct the final Story with answers to given questions as well as formulated questions.
- ☐ Present as a Level III synchronous formal presentation.

End Result

The end result of this project will be a **completed Tableau Story** and live presentation with an instructor in the **Level III formal synchronous format**!