Project 4

Caroline Badocha, Robbie Hockley & Angela Alexander Smith.

Project Proposal & Hypothesis:

Using the following dataset we [🎹 Spotify Tracks Dataset | Kaggle](https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset) we will clean the data and aim to answer the following questions and provide a range of visualizations:

1. How does Genre impact popularity?
2. How does Length of Song impact popularity?
3. What other attributes impact a song's popularity?

We will also then utilize this data to complete predictions for the future.

We plan to use Jupyter notebooks for coding and import the following libraries

* **import** pandas **as** pd
* **import** numpy **as** np
* **import** matplotlib.pyplot **as** plt
* **import** seaborn **as** sns
* **import** plotly.express **as** px
* **import** plotly.graph\_objects **as** go
* **from** plotly.subplots **import** make\_subplots
* **import** os
* **import** random
* **from** contextlib **import** contextmanager
* **from** time **import** time
* **from** tqdm **import** tqdm
* **import** lightgbm **as** lgbm
* *# import category\_encoders as ce*
* **from** sklearn.metrics **import** classification\_report, log\_loss, accuracy\_score
* **from** sklearn.metrics **import** mean\_squared\_error
* **from** sklearn.model\_selection **import** KFold
* **from** sklearn.model\_selection **import** train\_test\_split
* **from** sklearn.linear\_model **import** LinearRegression
* **from** sklearn.preprocessing **import** LabelEncoder
* **from** sklearn.preprocessing **import** StandardScaler
* **from** sklearn.decomposition **import** PCA
* **from** sklearn.manifold **import** TSNE
* **from** sklearn.cluster **import** KMeans
* **import** warnings
* warnings**.**filterwarnings('ignore')

Further to this we would then be hosting these visualizations on Tableau Public