From Billboards to Grammy

Team Five

HAYLEY GIMBEL

NetSuite Consultant at NAPPJO

Joined Data Analytics Course to gain understanding of basic coding & data analytics concepts

JENNIFER RATLIFF

Vice President Finance
National Home Builder

Joined Data Analytics Course for a new challenge

MAYA JEFFERSON

2020 University of Texas B.S.A. Mathematics Graduate

Joined Data Analytics Course to learn basic coding and jump start career.

LIJOY JOMOL

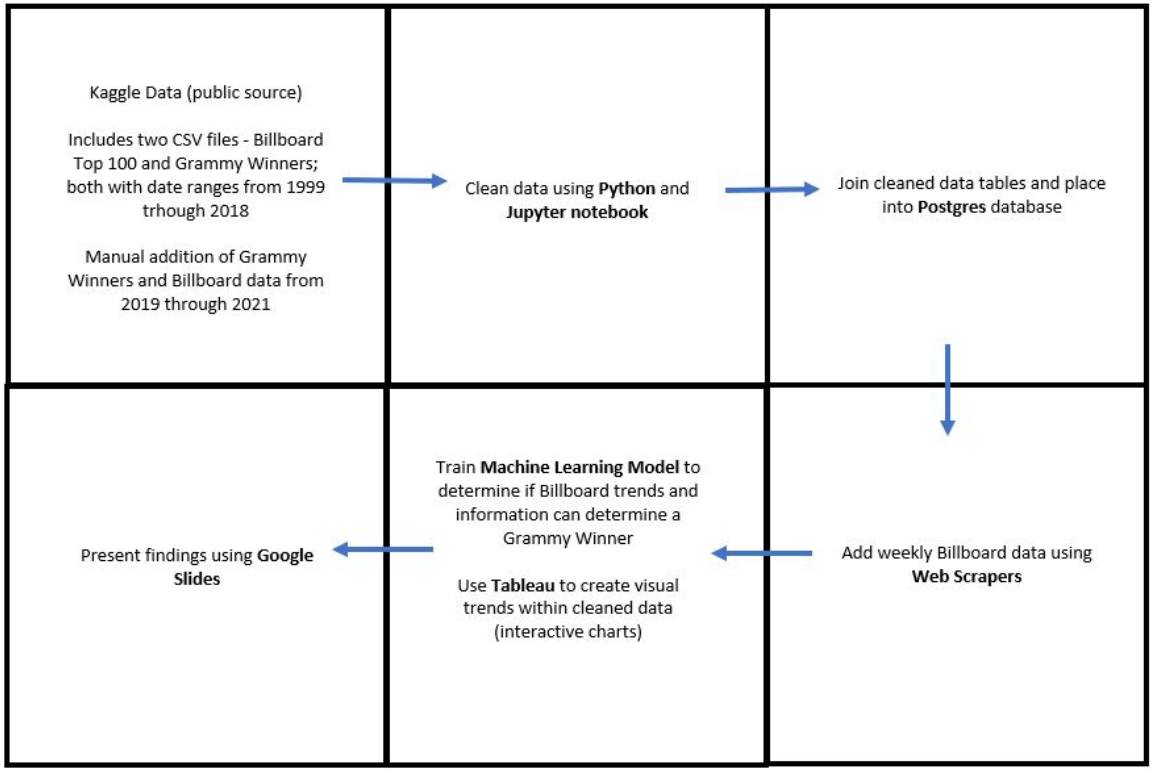
Certified Math Teacher

Joined Data Analytics Course for a career change.

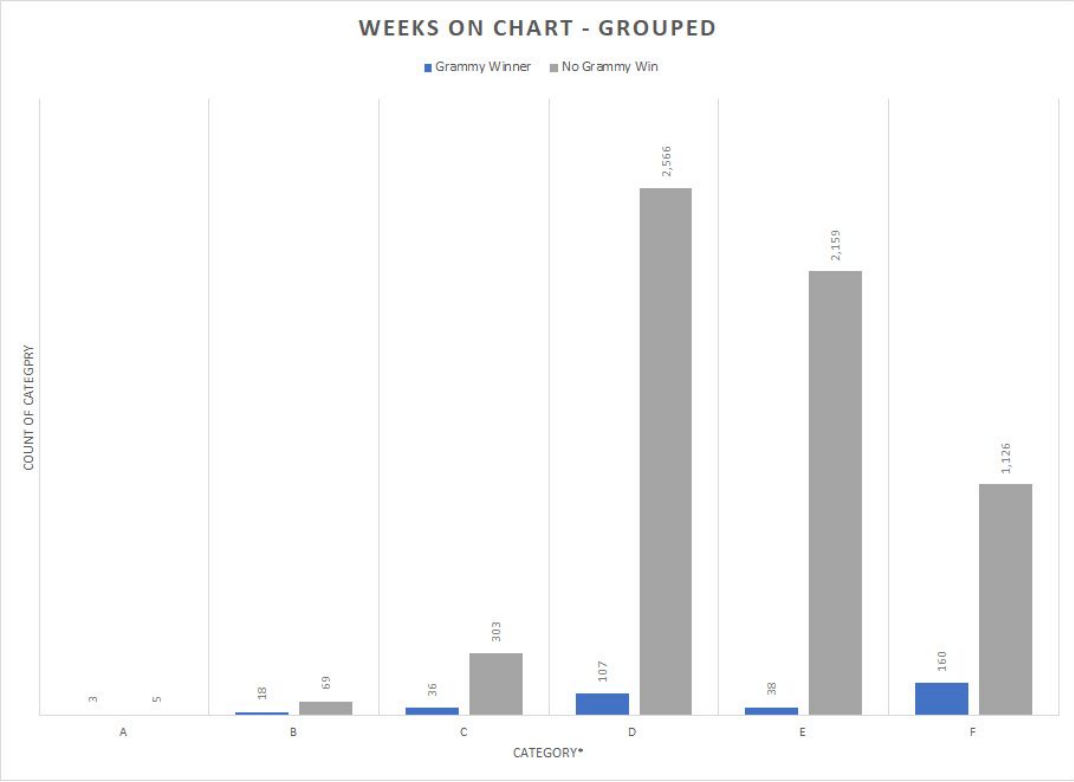
Purpose

- Determine if machine learning can identify Grammy winners based on
  - Weeks on Billboard Chart
  - Peak position on Billboard Chart
  - Age of song at time of nomination
- Selected topic due to team member work in music industry and data selection timeframe was during 2021 Grammy weekend
- Kaggle data included information for both Grammy winners and Billboard Top 100 from 1999 through 2018
  - Selected information from eight available tables
  - Specific tables chosen based on how final winners are selected (committee based)
- Data exploration - Team used basic Excel analysis to understand general information and determine best path forward
- Data Analysis - Team used Python with Jupyter notebook to clean files, combine tables, create basic charts, and determine options for machine learning

Project Dashboard

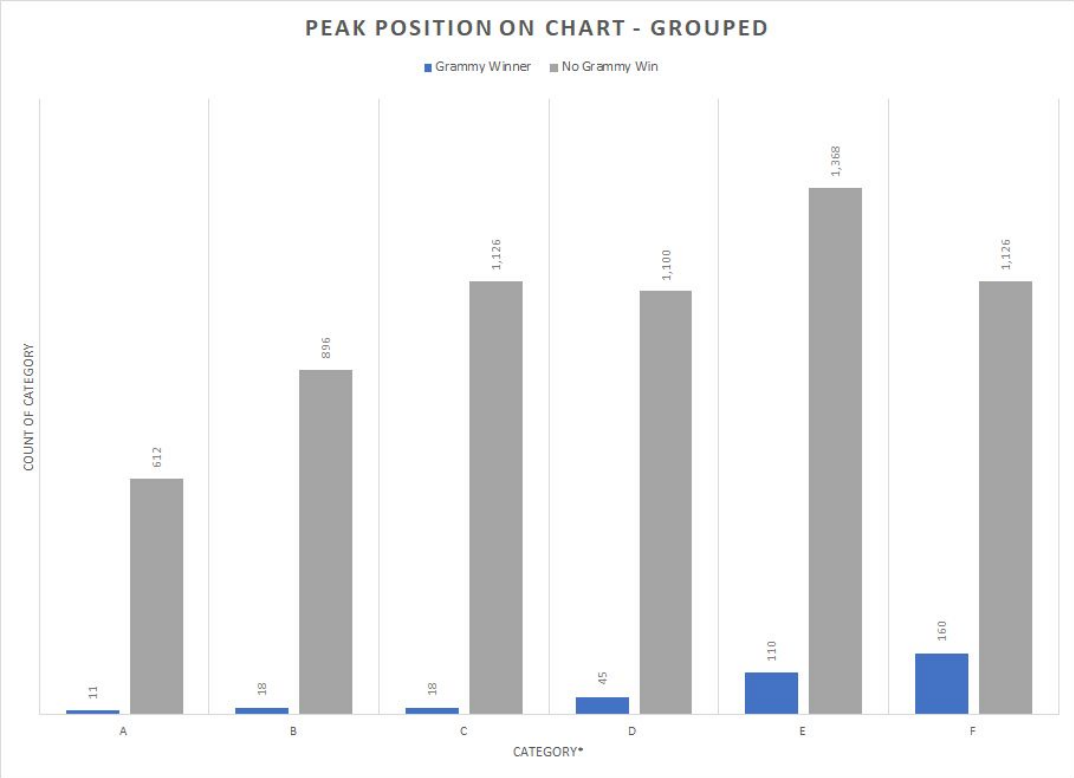


Data Exploration



- Completed simple statistical analysis on weeks on chart, such as minimum weeks, maximum weeks, and average weeks on chart.
- Grouped categories based on statistical results
    - A - More than 60 weeks
    - B - More than 45 weeks
    - C - More than 30 weeks
    - D - More than 15 weeks
    - E - Less than 15 weeks
    - F - Not ranked*
- Initial results showed little correlation between weeks on chart and a Grammy win (r = 0.008)

Data Exploration



- Completed simple statistical analysis on weeks on chart, such as minimum weeks, maximum weeks, and average weeks on chart.
- Grouped categories based on simple buckets
    - A - Greater than 80
    - B - Greater than 60
    - C - Greater than 40
    - D - Greater than 20
    - E - Less than 20
    - F - Not ranked*
- Initial results showed some correlation between peak position on chart and a Grammy win (r = 0.14)

Machine Learning - Grammy Prediction

The Machine learning model will provide data and charts once complete.

Model is still in training and final results are not available at this time.

Other factors (null hypothesis)

- What other factors could contribute to winning a Grammy?
  - Genre
  - Release date of song (summer, winter)
  - Multiple artists
    - Combination of well known and lesser known

Next Steps

How to improve the model
Additional features to include

Dashboard Link

https://public.tableau.com/profile/jomol#!/vizhome/BillboardtoGrammyDashboard/Dashboard2?publish=yes