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.
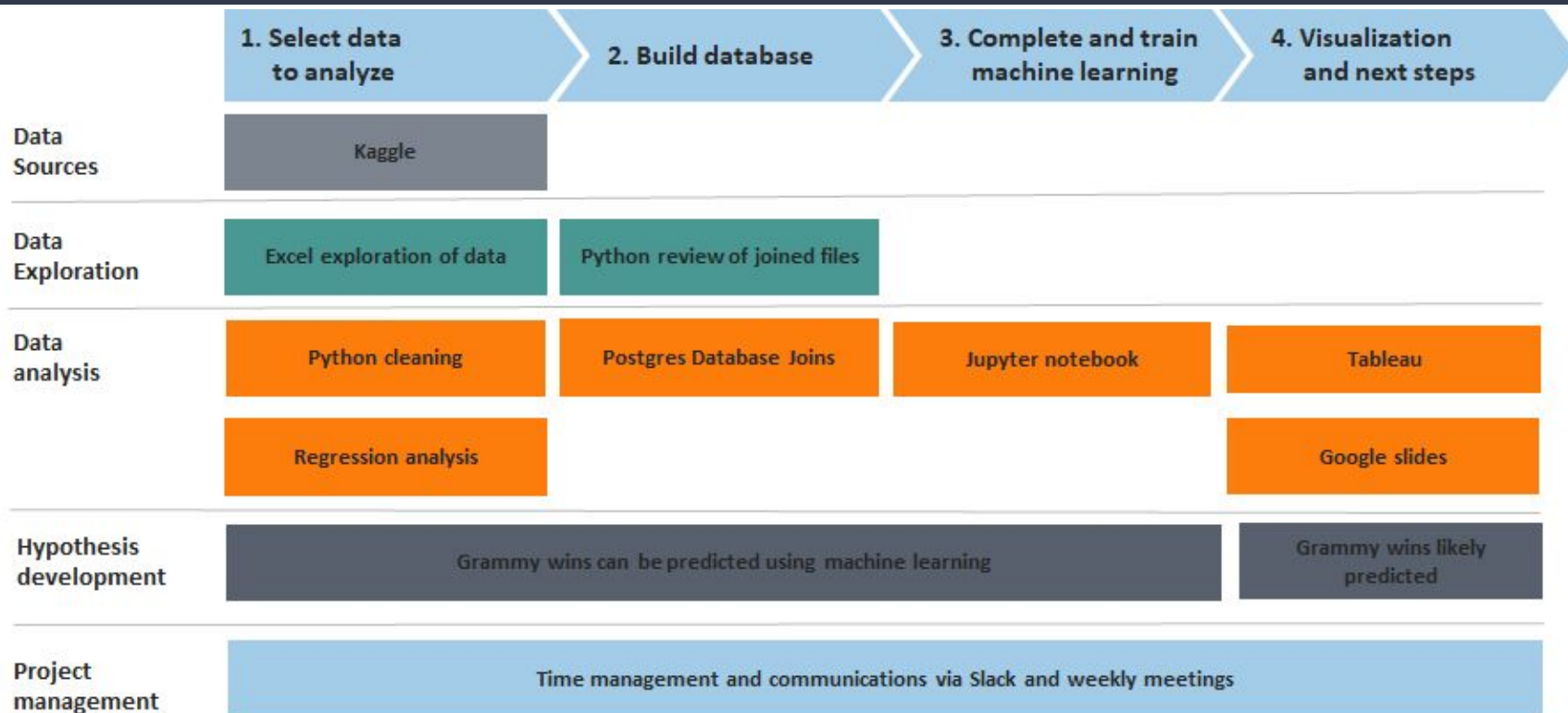
LIJOY JOMOL

Certified Math Teacher

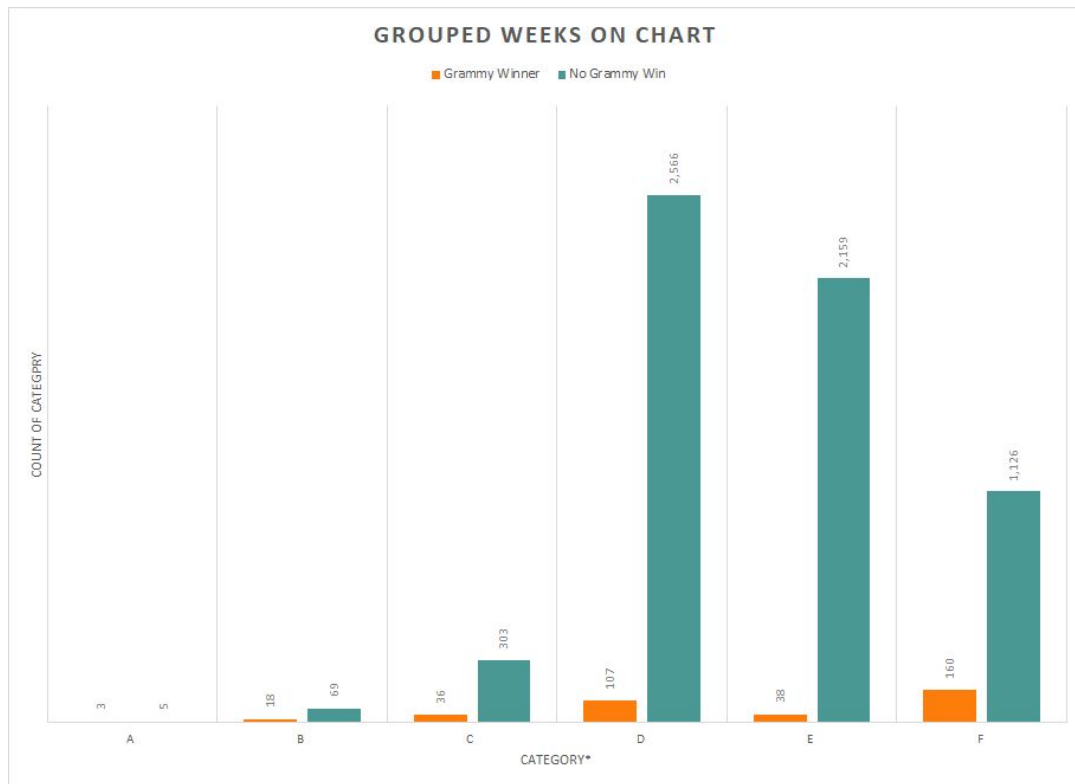Joined Data Analytics Course for a career change.

Purpose

- Determine if machine learning can identify Grammy winners based on
  - Weeks on Billboard Chart
  - Peak position on Billboard Chart
  - Genre
- 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 Timeline

	1. Select data to analyze	2. Build database	3. Complete and train machine learning	4. Visualization and next steps
Data Sources	Kaggle			
Data Exploration	Excel exploration of data	Python review of joined files		
Data analysis	Python cleaning	Postgres Database Joins	Jupyter notebook	Tableau
	Regression analysis			Google slides
Hypothesis development	Grammy wins can be predicted using machine learning			Grammy wins likely predicted
Project management	Time management and communications via Slack and weekly meetings			

Data Exploration

GROUPED WEEKS ON CHART

■ Grammy Winner ■ No Grammy Win

COUNT OF CATEGPRY

A: 3, 5
B: 18, 69
C: 36, 303
D: 107, 2,566
E: 38, 2,159
F: 160, 1,126

CATEGORY*

- Total weeks on Billboard Chart
- 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
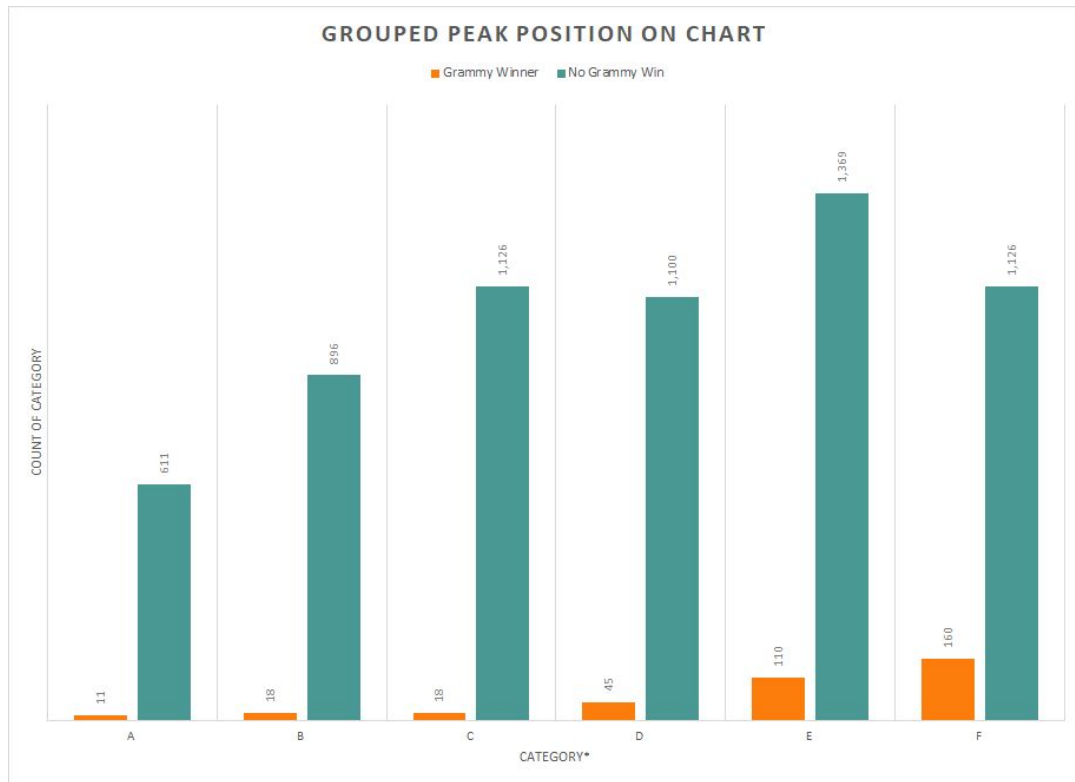
GROUPED PEAK POSITION ON CHART

- Peak Position on Billboard Chart
- 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.

The Machine Learning Model
- Logistic Regression
  - Predicts binary outcomes
  - Whether or not an artist that is on the Billboard Top 100 Chart will win a Grammy

Training and Testing:

- Use the train_test_split module to split X and y into training and testing sets: X_train, X_test, y_train, y_test.
- y_test (whether or not an artist that is on the Billboard Top 100 Chart will win a Grammy) are the outcomes from the original dataset that were set aside for testing.
- The model's predictions, y_pred, were compared with the actual values, y_test.

Accuracy:
The MLM will accurately predict whether an artist will receive a grammy based on their Billboard Top 100 Chart performance 94.12% of the time.

Dashboard Link

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

Other factors and considerations

- Dataset limitations - time on chart and peak position unavailable
- Grammy Award Categories
  - Additional categories throughout years
- Committee based decision bias
- Multiple genre listed
- Additional information for MLM
  - Genre impact
  - Release date in relation to Grammy awards
  - Location of release (US, UK)

XXXXXXXX

Next Steps and Additions

- Automatic web scraping to add in weekly Billboard chart
- Use additional data sets and modern compilations of song ranking
- Expand Grammy and Billboard datasets to include all available
- Grammy Award category changes and additions
  - 1959 - 14 categories
  - 2021 - 84 categories

Questions?