From the Billboard Hot 100 to the Grammy Awards

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.

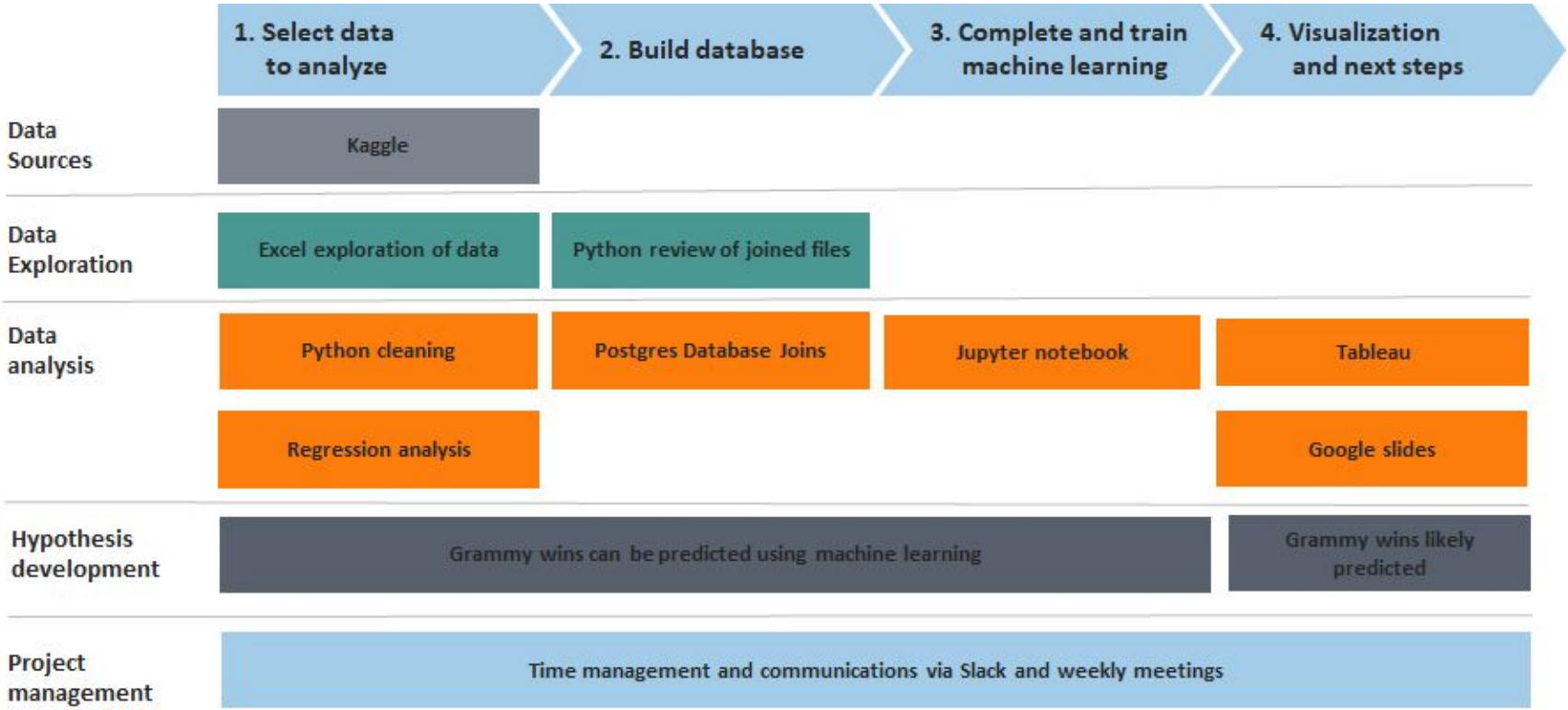
Question

Does inclusion on the Billboard Hot 100 Chart impact the likelihood that a song will win a Grammy Award?

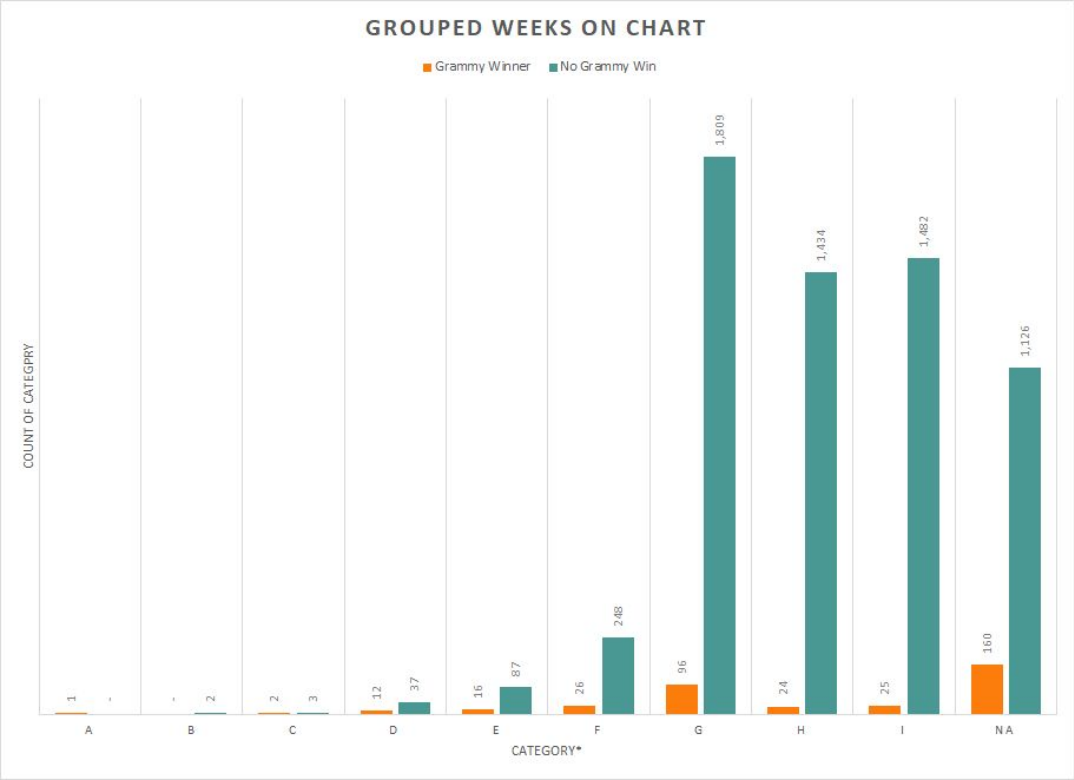
Anderson Paak & Bruno Mars -
“Sonic Song” currently #1 on Billboard Hot 100

Taylor Swift -
11- time Grammy Award Winner

Project Outline

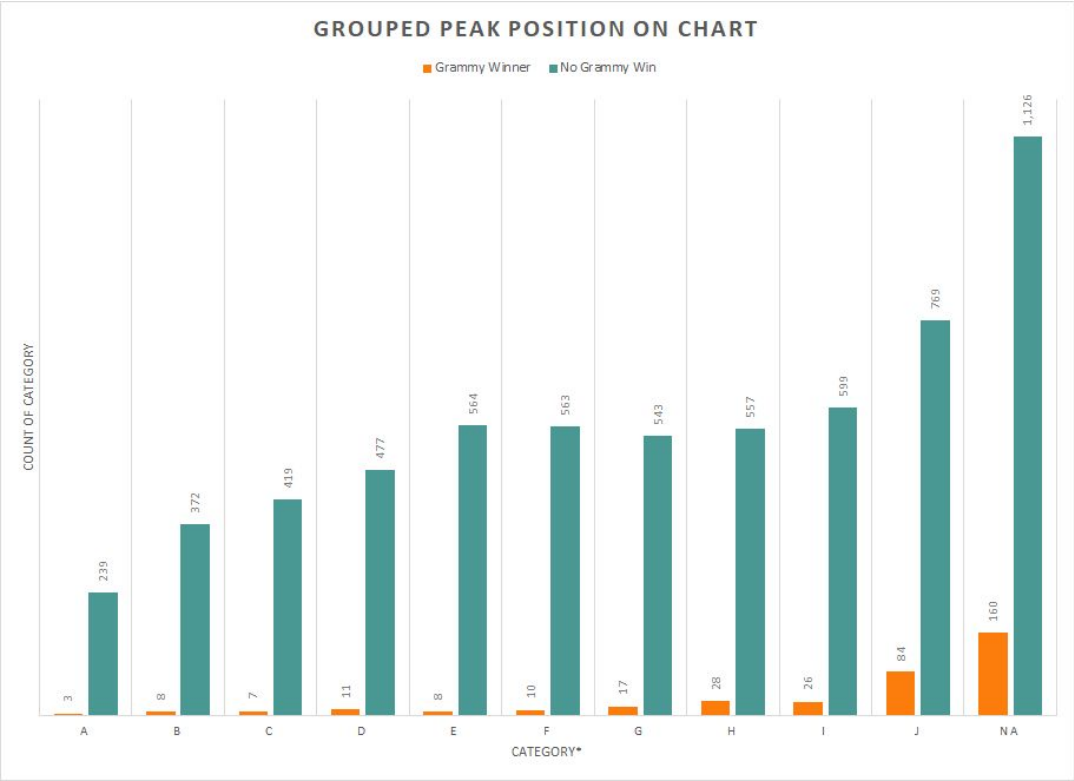


Data Exploration



- Tools Used: Excel
- Total weeks on Billboard Chart
- Grouped categories based on statistical results
    - A - Greater than or equal to 80
    - B - Greater than or equal to 70
    - C - Greater than or equal to 60
    - D - Greater than or equal to 50
    - E - Greater than or equal to 40
    - F - Greater than or equal to 30
    - G - Greater than or equal to 20
    - H - Greater than or equal to 10
    - I - Less than 10
    - NA - Not ranked
- Initial results showed little correlation between weeks on chart and a Grammy win (r = 0.008)

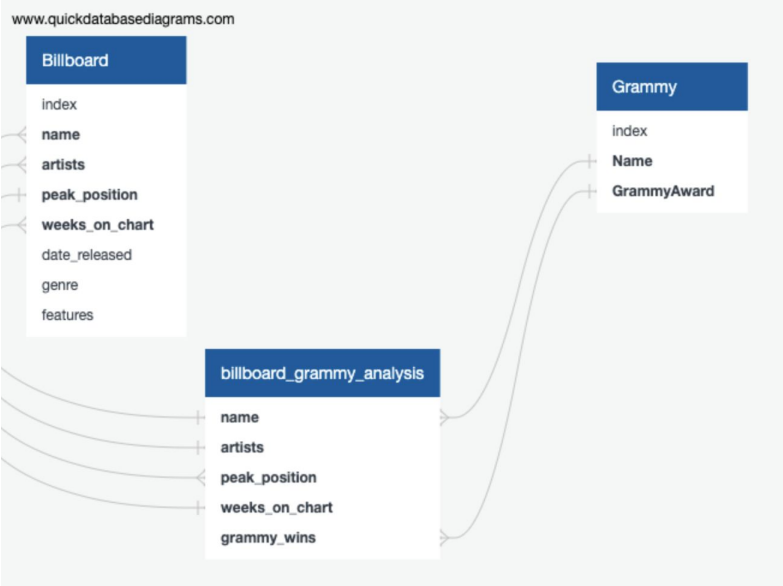
Data Exploration



- Peak Position on Billboard Chart
- Grouped categories based on simple buckets
  - A - Greater than or equal to 90
  - B - Greater than or equal to 80
  - C - Greater than or equal to 70
  - D - Greater than or equal to 60
  - E - Greater than or equal to 50
  - F - Greater than or equal to 40
  - G - Greater than or equal to 30
  - H - Greater than or equal to 20
  - I - Greater than or equal to 10
  - J - Less than 10
  - NA - Not ranked
- Initial results showed some correlation between peak position on chart and a Grammy win (r = 0.14)

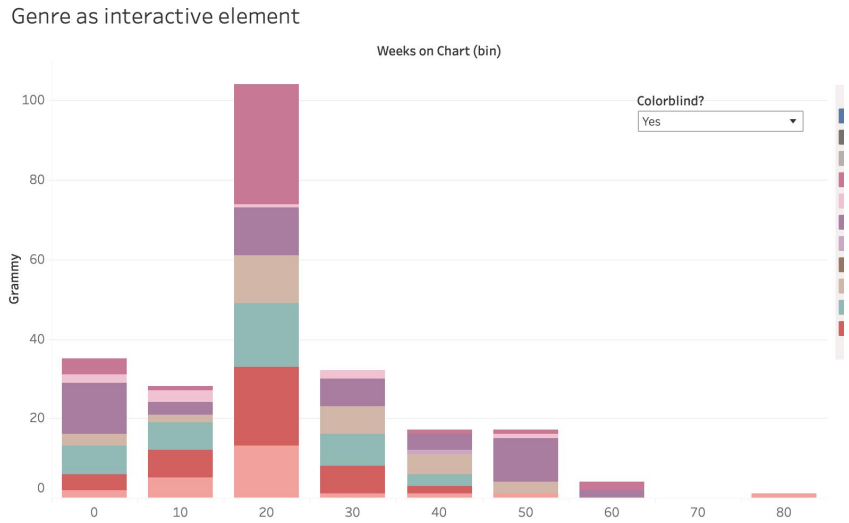
Database Design

- Created ERD
- Finalized Postgres as the database.
- Used Python to clean Billboard and Grammy CSVs
- Loaded CSVs into Postgres database tables
- Joined Billboard and Grammy tables to get one complete dataset.



Used an iterative design approach

Tableau Dashboard



- Tool Used: Tableau
- Dashboard
- The interactive chart shows the correlation between weeks on chart and Grammy win for a specific Genre. We can either look at a specific Genre or the consolidated view.
- Also, to help user who are color blind, we have added the color-blind view of the chart.
- Interactive features include genre filters

Machine Learning - Grammy Prediction

	Prediction	Actual
0	1	1
1	2	2
2	1	1
3	2	2
4	1	1
5	1	1
6	1	1
7	1	1
8	2	2
9	1	1
10	2	2

Machine Learning Model Choice

- Logistic Regression
  - Predicts binary outcomes
  - Used to help assess whether or not an artist that is on the Billboard Top 100 Chart will win a Grammy
- Benefits:
  - Simple to implement and make predictions for linear outcomes
  - Simple to understand, train, and update with new data to be used in the future
- Limitations:
  - Does not work with non-linear outcomes
  - Requires a large dataset
  - Prone to overfitting

Machine Learning - Grammy Prediction

Machine Learning Model In Action

- Feature Selection:
  - Dropped Variables: "weekly_rank", "writing_credits", "lyrics"
  - These variables are not informative for our MLM analysis.
  - Increased Accuracy Score
- Training and Testing:
  - X predicts y
  - X: input created by dropping “artists” and “name” columns
  - y: output taken from “GrammyAward” column
  - Use the train_test_split module to split X and y into training and testing sets
    - X_train, X_test, y_train, y_test
  - The model’s prediction, y_pred, were compared with the actual values, y_test.
- Accuracy Score:
  - Assesses how well our Machine Learning Model Performs
  - **94.12%**

Results and Other Considerations

- **Original hypothesis:** Grammy winners can be accurately predicted based on Billboard Chart information
- **Null hypothesis:** Grammy wins do not follow Billboard chart positions
- Final result of machine learning model showed a 94.12% accuracy score
    - Conclusion from model that Grammy winners are likely predicted using current model.
- Dataset limitations - time on chart and peak position unavailable
- Committee based decision bias to data analytics based
- Multiple genre listed for individual song
- Additional information for later machine learning models
    - Genre impact
    - Release date in relation to Grammy awards
    - Location of release (US, UK)

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?