# Musical Insights

Kathrin and Kevin

# Problems to solve

1. Have Billboard 100 hit songs changed from the 2000's to 2010's in their song length?

2. Does vulgarity have an effect on a song's popularity?

3. Does the tempo of a song affect its peak rank on the Billboard 100?

4. Is there a relationship between the "musicality" of songs and how high they chart on the Billboard 100?

# Data

- Scraped songs from billboard top 100 for 2003-2018
  - Name, artist, chart ranking, duration on chart,  peak rank, etc

- Used Spotify API to gain additional info
  - Bpm, vulgarity, valence, duration,  etc.

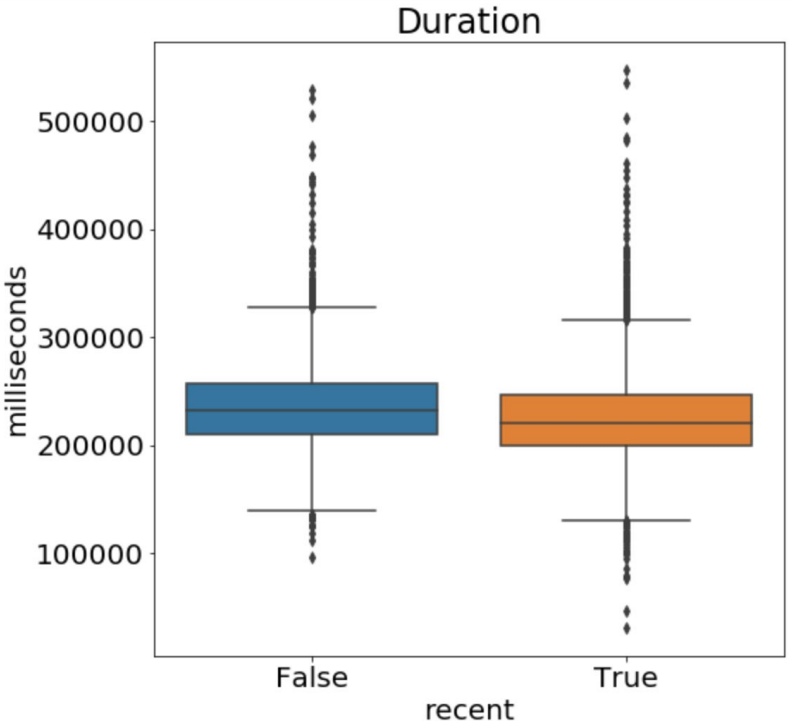
# 1 Changes in Song Length in Hit Songs from 2000's to 2010's

## Hypotheses:

- **Null:** No difference in length between recent and older songs
- **Alternative:** A difference exists in length between old and recent songs

## Chi-Squared Test

- **Independent Variable:** Whether song is recent (2010's) or old (2000's)
- **Dependent Variable:** Length of song in milliseconds




## Results:

- Test statistic = 130
- P-value = 5.0e-22
- Reject null hypothesis and conclude that there is a difference in length between recent and older songs

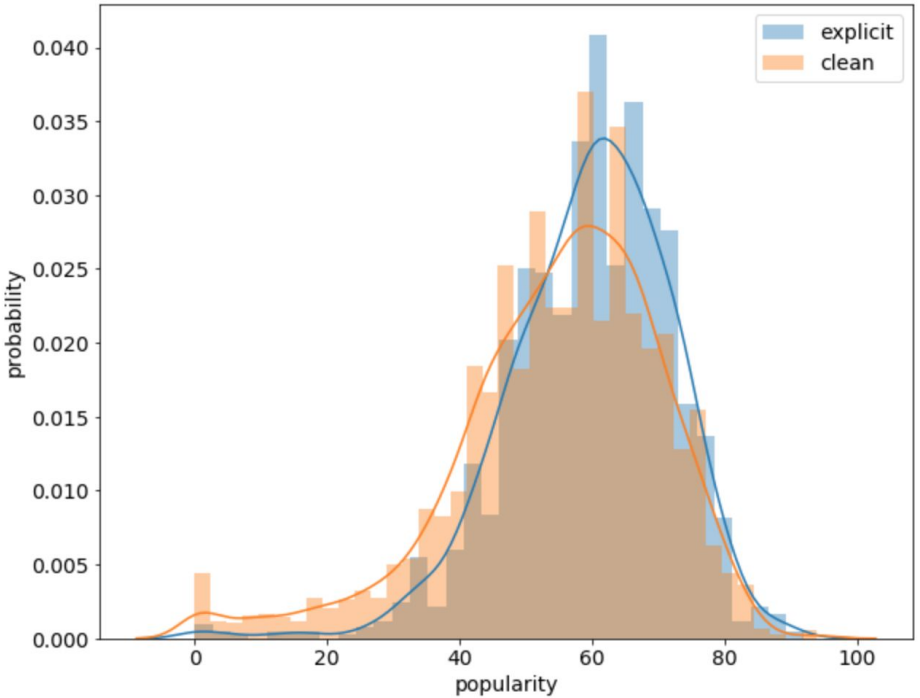
# 2 Changes in popularity in between explicit and clean songs

## Hypotheses:

- **Null:** No difference in popularity between clean and explicit songs
- **Alternative:** A difference exists in popularity between clean and explicit songs

## Two Sample T-Test

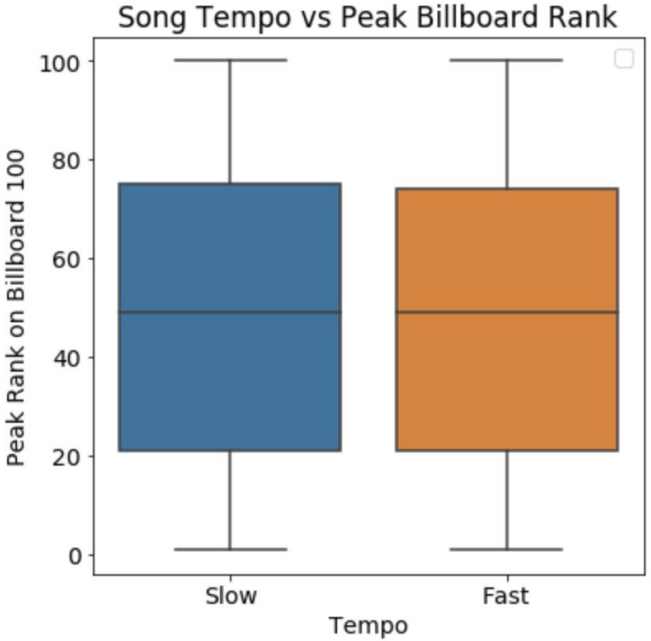
- **Independent Variable:** Whether song is explicit or clean
- **Dependent Variable:** Spotify popularity of song




## Results:

- Test statistic = 12.5
- P-value = 2.5e-35
- Reject null hypothesis and conclude that a difference exists in popularity between clean and explicit songs

# 3 Tempo of a Song vs Peak Chart Rank




## Hypotheses:

- **Null**: Songs with a fast tempo (> 120 bpm) do not reach a higher rank on the Billboard 100 charts than slow tempo songs (< 120 bpm)

- **Alternative**: There is a relationship between song tempo (fast vs slow) and how high it charts

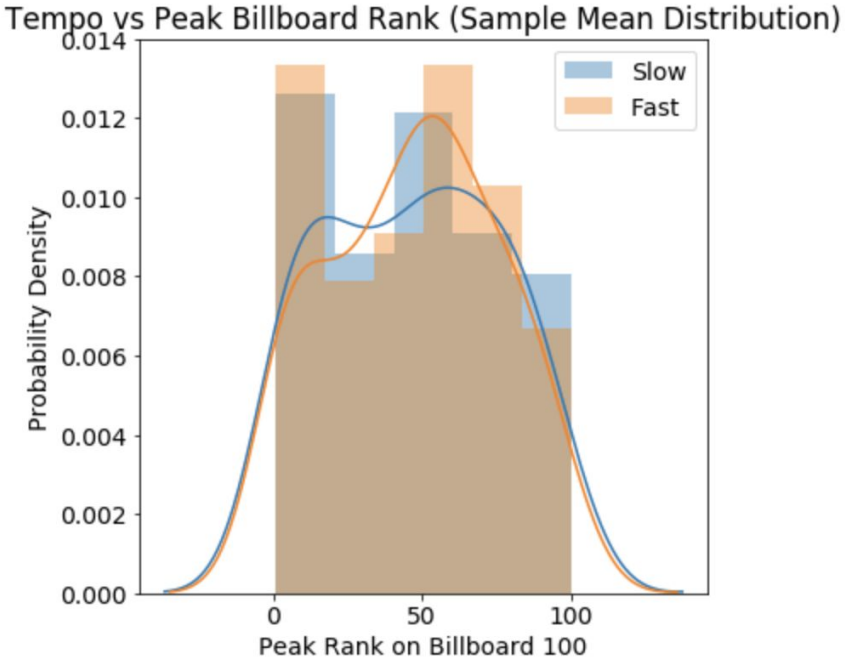
## 2 Sample T-Test

- **Dependent Variable**: Peak rank
  - **Note**: Peak rank was not normally distributed, so a mean sampling distribution was created (n=100)
- **Independent Variables**: Release Season (categorical)

# 3 Tempo of a Song vs Peak Chart Rank

## Conclusion

- **T-Statistic = 0.0586**
- **P-Value = 0.95322**
- **Accept the Null Hypothesis** - There is no difference in how high a slow song charts on the Billboard 100 compared to a fast song



# 4 Relationship btwn “musicality” of songs and peak chart rank on the Billboard 100

## ANOVA

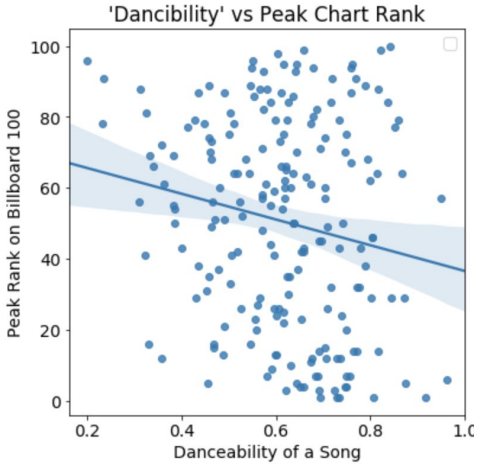
- Dependent Variable: Peak rank
- Independent Variables: Musicality
  - Danceability
  - Energy
  - Key
  - Loudness
  - Mode
  - Speechiness
  - Acousticness
  - Liveness
  - Valence

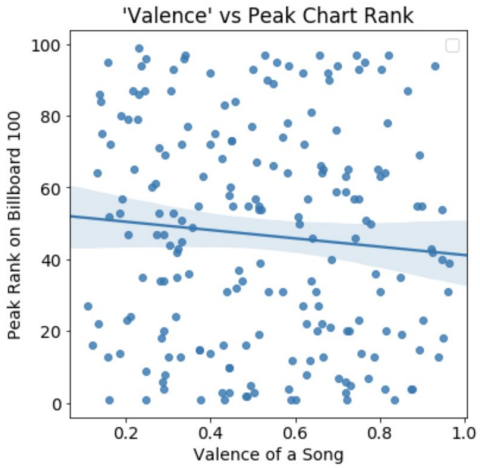
| | Sum of Squares | dof | F-Test Statistic | p-value |
|---|---|---|---|---|
| ★ danceability | 32011.721 | 1.0 | 36.033 | 0.000 |
| ★ energy | 23360.131 | 1.0 | 26.295 | 0.000 |
| key_ | 1347.310 | 1.0 | 1.517 | 0.218 ⊘ |
| ★ loudness | 5124.551 | 1.0 | 5.768 | 0.016 |
| mode | 512.385 | 1.0 | 0.577 | 0.448 ⊘ |
| speechiness | 607.476 | 1.0 | 0.684 | 0.408 ⊘ |
| ★ acousticness | 25704.644 | 1.0 | 28.934 | 0.000 |
| liveness | 17.857 | 1.0 | 0.020 | 0.887 ⊘ |
| ★ valence | 9366.742 | 1.0 | 10.544 | 0.001 |
| Residual | 5047821.815 | 5682.0 | NaN | NaN |

## Hypotheses:

- **Null**: There is **no relationship** between the “musicality” of a song and how high it charts
- **Alternative**: There **exists a relationship** between the musicality of a song and how high it charts

# 4 Relationship btwn “musicality” of songs and peak chart rank on the Billboard 100




OLS Regression Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | peak_rank | **R-squared:** | 0.023 |
| **Model:** | OLS | **Adj. R-squared:** | 0.021 |
| **Method:** | Least Squares | **F-statistic:** | 14.81 |
| **Date:** | Fri, 04 Oct 2019 | **Prob (F-statistic):** | 5.01e-24 |
| **Time:** | 09:44:25 | **Log-Likelihood:** | -27394. |
| **No. Observations:** | 5692 | **AIC:** | 5.481e+04 |
| **Df Residuals:** | 5682 | **BIC:** | 5.487e+04 |
| **Df Model:** | 9 | | |
| **Covariance Type:** | nonrobust | | |

## Conclusion:

- **While we rejected the null hypothesis that there is no relationship between the “musicality”** of a track and how high it reaches on the charts, **the size of the effect** (adjusted R-squared) **appears to be too small to take advantage of. Only 2 % of the peak chart rank can be explained by the “musicality” of a song.**

# Thank you.
# Any questions?