



# ANALYSING MUSIC REVIEWS

Lee Mackey | Anna Zubova

---

SQL, PostgreSQL, pandas  
Dataset: around 18,000  
reviews

T-test, z-test,  
classification  
algorithms, McNemar  
test

How can we apply our  
findings?



What do we want to  
know?

Based on statistical  
tests, what did we find?

---

# Questions

## 1. Are metal and jazz genres average scores different?

Methodology: 2-samples t-test to compare 2 groups

Findings: Yes, there is statistical difference in metal and jazz scores

Applications: Test all genres, possibly create an algorithm to compensate for lower scores in some genres

---

---

# Questions

## 2. Multi-genre reviewers: do the same reviewers score metal and jazz genres differently?

Methodology:	Paired t-test across multi-genre reviewers, for jazz and metal
Findings:	No average difference found. However, may need more test data.
Applications:	Understand same reviewers' behaviour across different genres

---

---

# Questions

## 3. Scores produced by each author: different from the average?

Methodology:	Z-test, random samples from each author's review scores (where more than 50)
Findings:	Yes, around 40% of the reviewers have average score different from total average
Applications:	Vet reviewers that produce very high/low average scores

---

---

# Questions

## 4. Compare two text-based algorithms in predicting review genres?

Methodology:	Compare two algorithms' agreement on rock vs non-rock class using evaluation metric (McNemar statistic)
Findings:	Yes. but technical details & just this study, reviews, metric?
Applications:	Improve internal choices between competing classification algorithms through a metric portfolio approach

---

---

---

# Next steps

Understand the review content that users engage with.

Building out rigorous framework for evaluating reviewers/reviews across platforms

- Genres
  - Reviewers and Scores
  - Review texts
-

---

# Questions

---