
# Guide to Vinyl Noobs

Tim Miller

# VATICAN VAMPS

# Vinyl Me, Please is a record of the month club. The best damn record club out there, in fact.

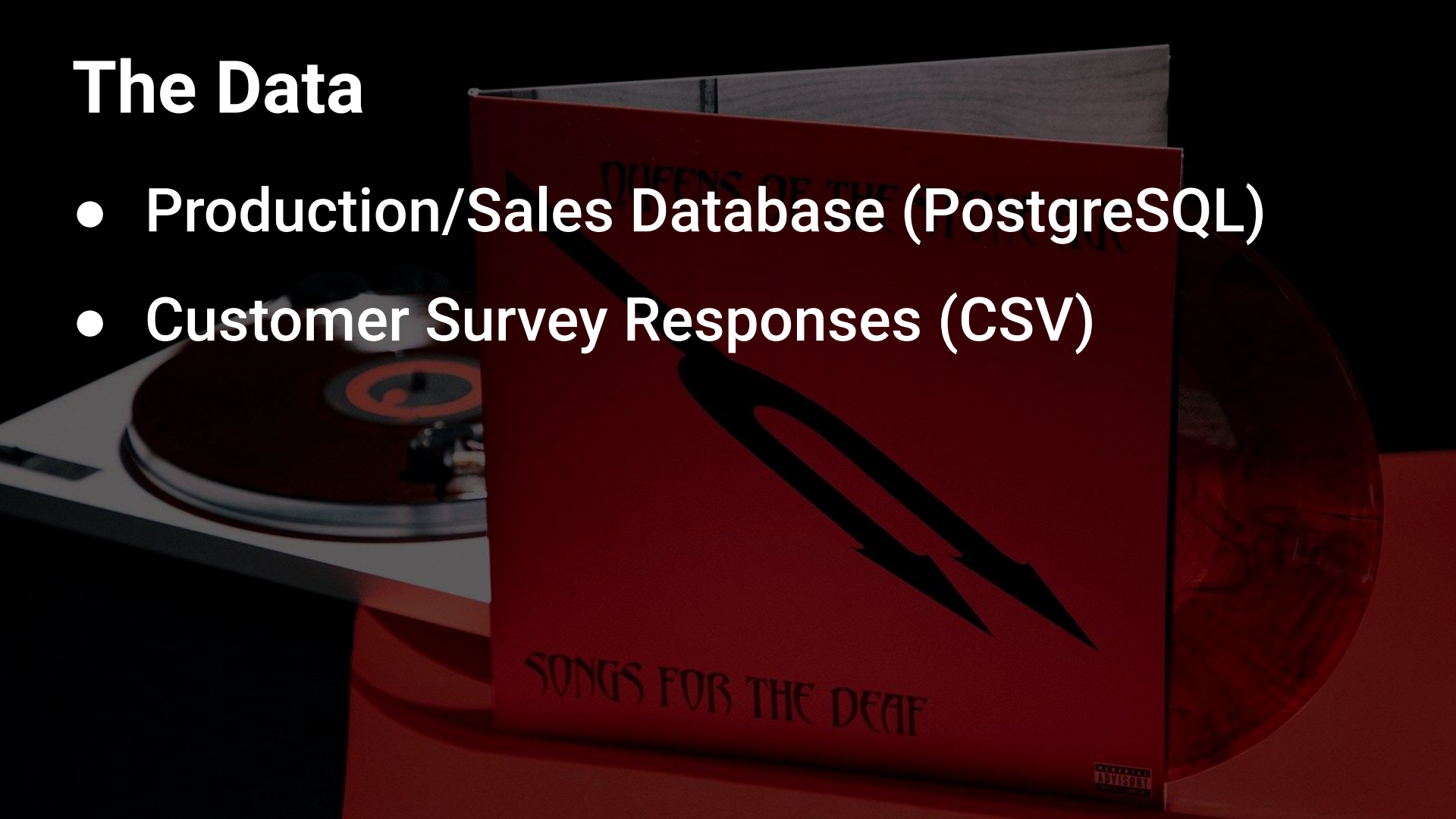
- New album releases & re-issues
- Subscription model


weezer
Pinkerton
on
off
THIRD MAN
MARCO PRICE
Small Black
Despicable Dogs
(Washed Out Rem


QUEENS OF THE STONE AGE
SONGS FOR THE DEAF
PARENTAL ADVISORY

# The Data

- Production/Sales Database (PostgreSQL)
- Customer Survey Responses (CSV)



# Initial Findings:

- Most valuable customers are long-time vinyl buyers (15+ years).
- Problem: Limited market!

Image credit: https://medium.com/cuepoint/the-secret-lives-of-vinyl-hoarders-fa519eee6fff

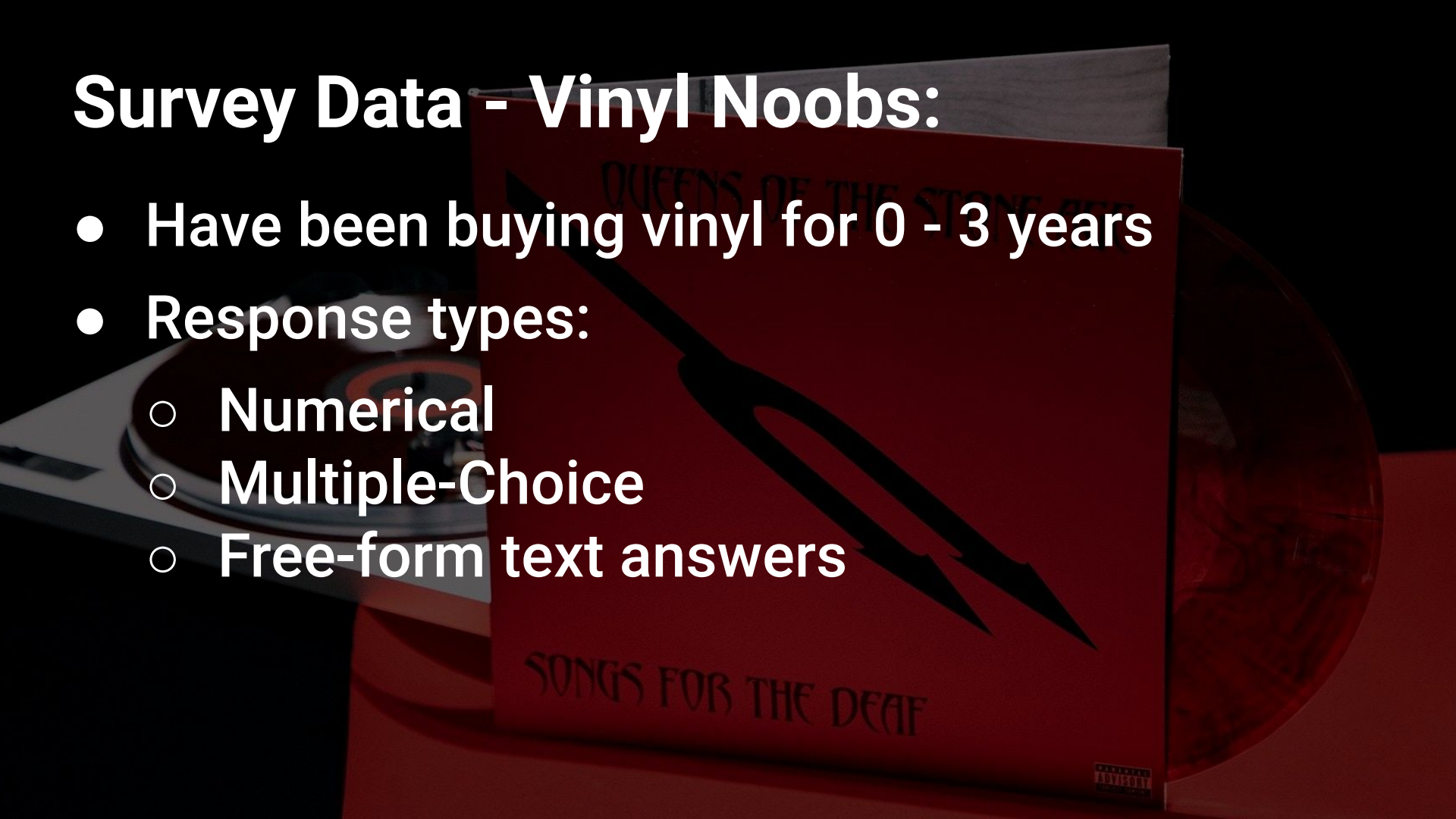
# Goal:

Investigate how to:

- Attract NEW vinyl buyers
- Retain them as subscribers

# Survey Data - Vinyl Noobs:

- Have been buying vinyl for 0 - 3 years
- Response types:
  - Numerical
  - Multiple-Choice
  - Free-form text answers



# How Does VMP Attract Vinyl Noobs?

Q: Why did you start buying vinyl originally?

Natural Language Processing - LDA

- Latent Dirichlet Allocation (LDA) assumes that each “document” (response) is generated from only a small set of topics, and that each topic is generated from a small set of words.

# Q: Why did you start buying vinyl originally?

## SKLearn LDA Topic Modeling with 5 topics for “Noobs”:

| | |
|---|---|
| **Topic 1: Gifts** | player, **record player**, record, got, **gift**, turntable, got record, **got record player, gifted**, received, **birthday**, way listen, **christmas**, gave, given |
| **Topic 2: Nostalgic Hipsters** | 2017, wanted support, **wanted support artists**, support artists, downhill, display, christmas, huge, felt, **childhood, beautiful, form music**, turntable christmas, **memories, passionate** |

# How Does VMP Retain Vinyl Noobs?

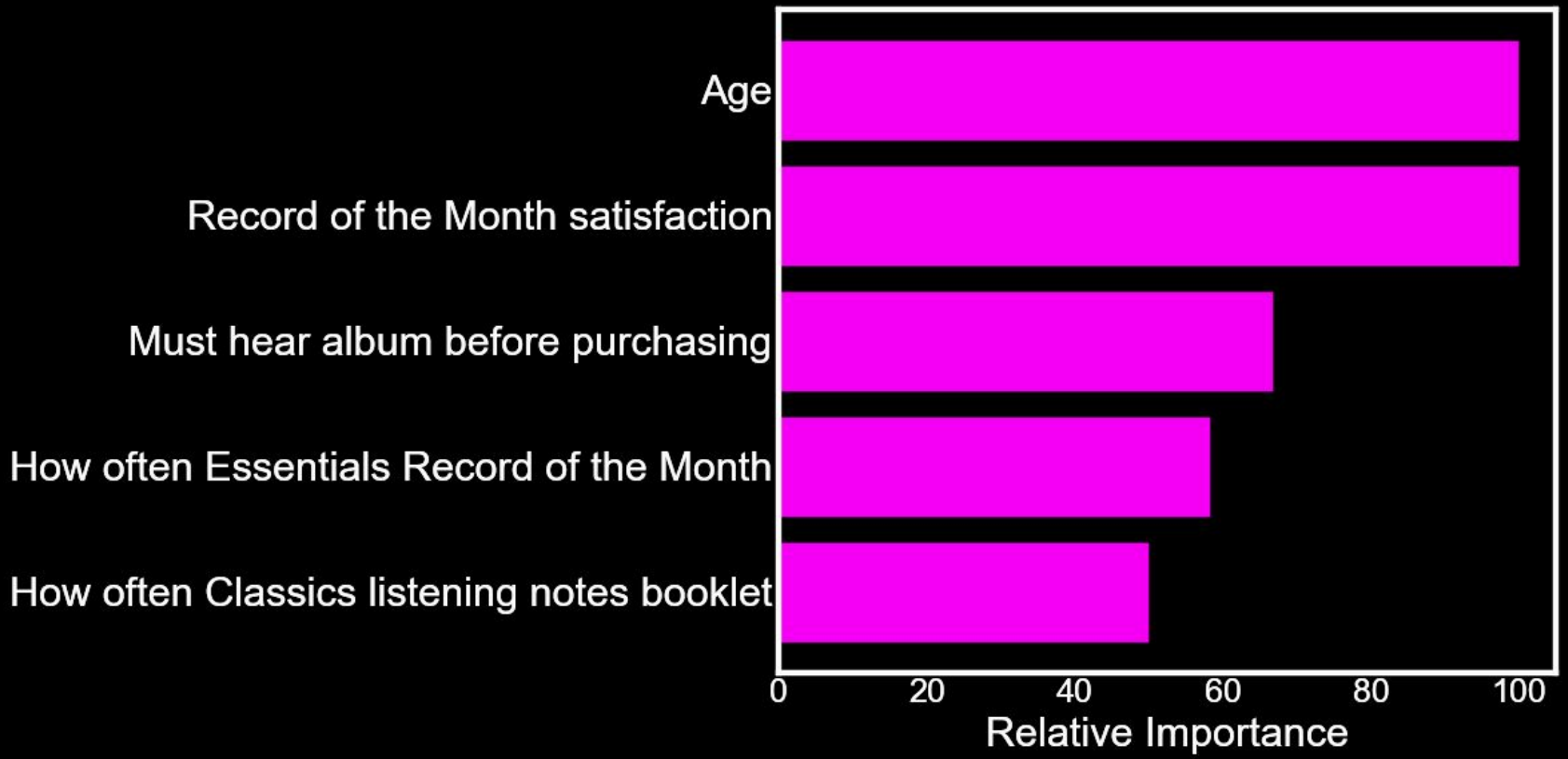
- Numerical/Multiple-Choice data
- Added account status from SQL database

# Model Selection

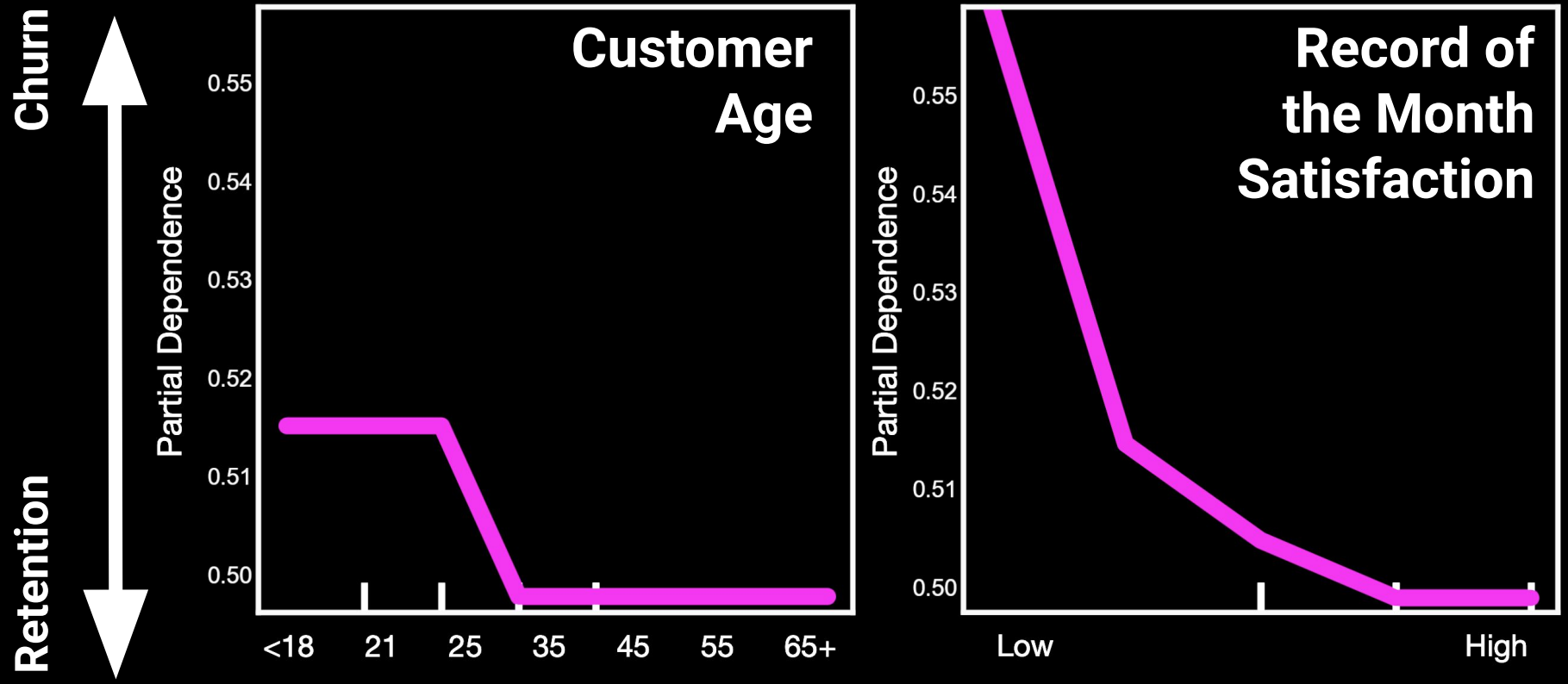
- Predicting Account Status (Active/Canceled)
- Large number of features (113 columns)

→ AdaBoost (Adaptive Boosting) Classifier

# AdaBoost Top Features

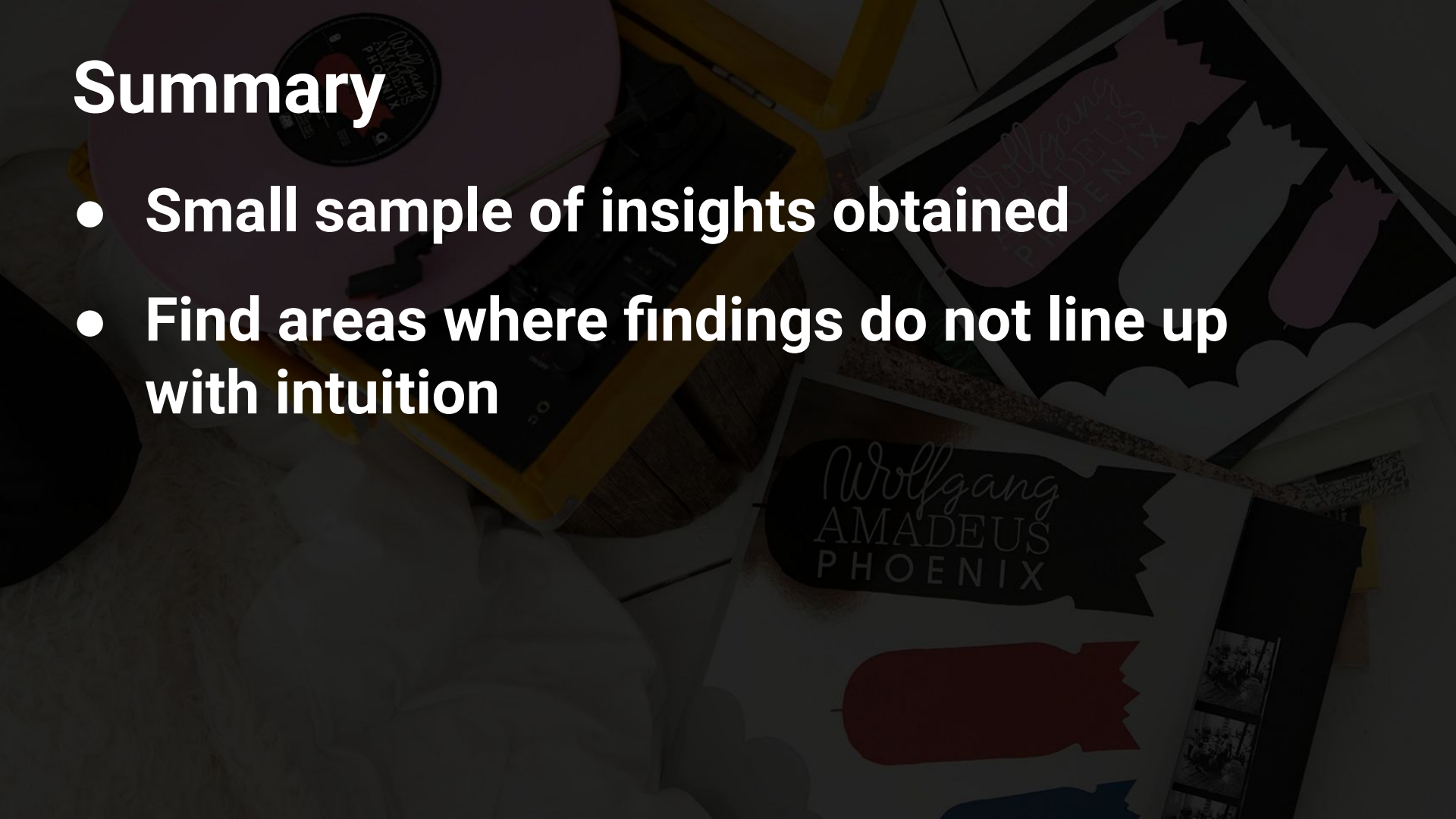


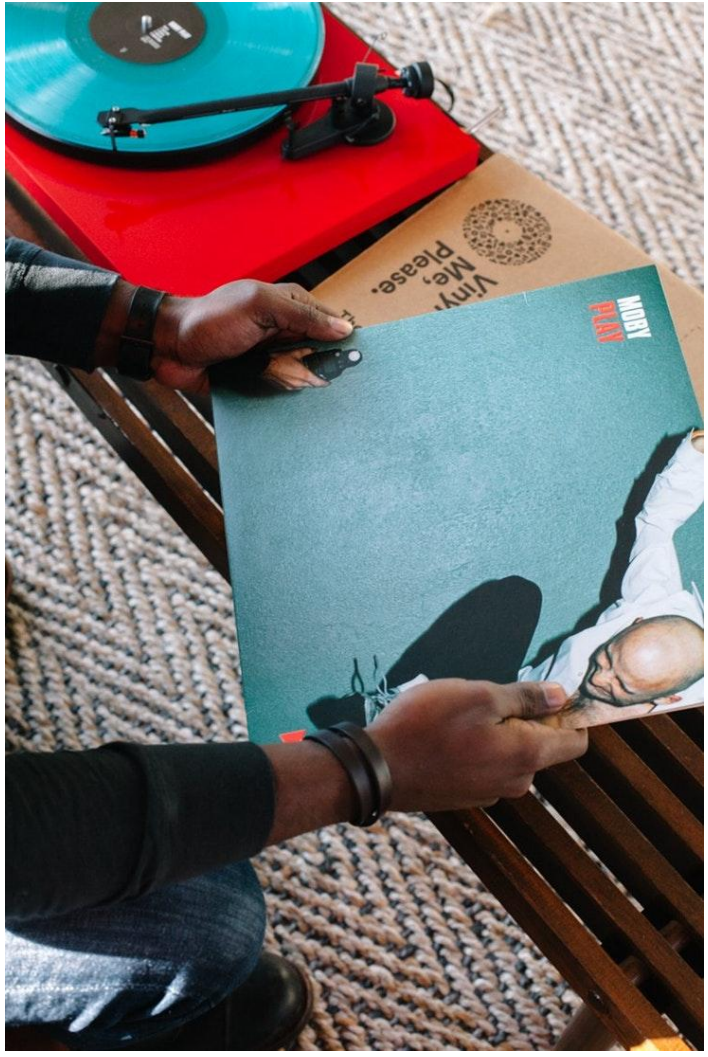
# AdaBoost Partial Dependence Plots



# Summary

- Small sample of insights obtained
- Find areas where findings do not line up with intuition






# Thank You!


## Tim Miller

Vinyl Me, Please.

tpmiller85@gmail.com
GitHub: /tpmiller85
LinkedIn: /tpmiller85


---

python™ PostgreSQL Psycopg2 pandas

$y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}$

NumPy

scikit learn ANACONDA matplotlib

# Create_models.py (class)

- Makes train/test split, performs grid search, pickles and saves models. Model options:
  - GradientBoostingClassifier
  - AdaBoostClassifier

# survey_nlp.py (class)

- Reads in raw text data for one open-ended Vinyl Me, Please customer survey question and performs various types of NLP. The following options are available:
  - SKLearn LDA
  - SKLearn NMF
  - Gensim LDA
  - Mallet LDA using Gensim wrapper.

For more info on Mallet LDA visit the homepage: http://mallet.cs.umass.edu