

Capstone Project Sprint 2

**Vancouver Travel Itinerary
Recommendation**

Izabel Lopez



TABLE OF CONTENTS



01.

INTRODUCTION

Business Problem,
Proposed Solution,
Impact

02.

DATA SCIENCE

Data Collection, Data
Cleaning,
Preprocessing, and
Baseline Model

03.

NEXT STEPS

Feature Engineering,
Hyperparameter
Tuning, Content-Based
Filtering, Scheduling



INTRODUCTION

- **Business Problem**
 - Travel planning is time consuming
 - Sightseeing Tours aren't flexible
- **Proposed Solution**
 - Recommendation system based on user input
- **Impact**
 - Saves Research Time
 - Itinerary Blueprint



DATA SCIENCE



DATA COLLECTION AND CLEANING

Expanded Geographical Scope, Collected Geographical Details, Reduce Main Categories



PREPROCESSING

Location, Number of Reviews, Average Rating, Main Category , Review Keywords



BASELINE MODEL

Various Clustering Models evaluated by Silhouette Score



DATA COLLECTION AND CLEANING

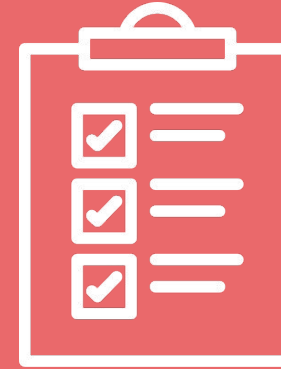


Expanded Geographical Scope



Burnaby, Richmond,
West Vancouver,
North Vancouver

Collected Geographical Details



Latitude, Longitude

Reduce Main Categories



MC_park, MC_restaurant,
MC_shopping, MC_tourist

PREPROCESSING



LATITUDE AND LONGITUDE

Distribution and Log Transformation



REVIEW KEYWORDS

CountVectorizer and Text Analysis



REVIEWS AND RATINGS

Distribution and Log Transformation

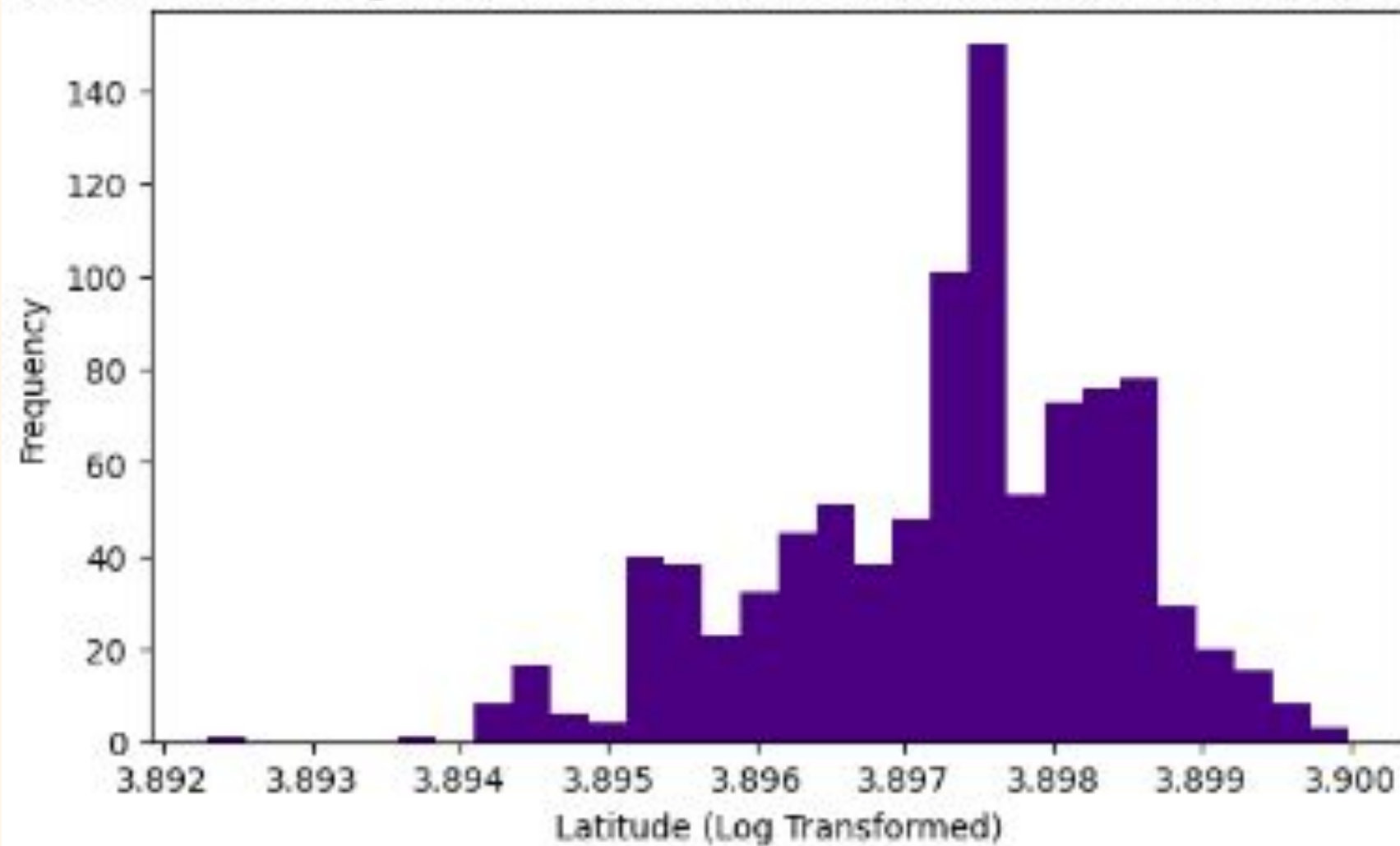


FINAL MODEL DATA

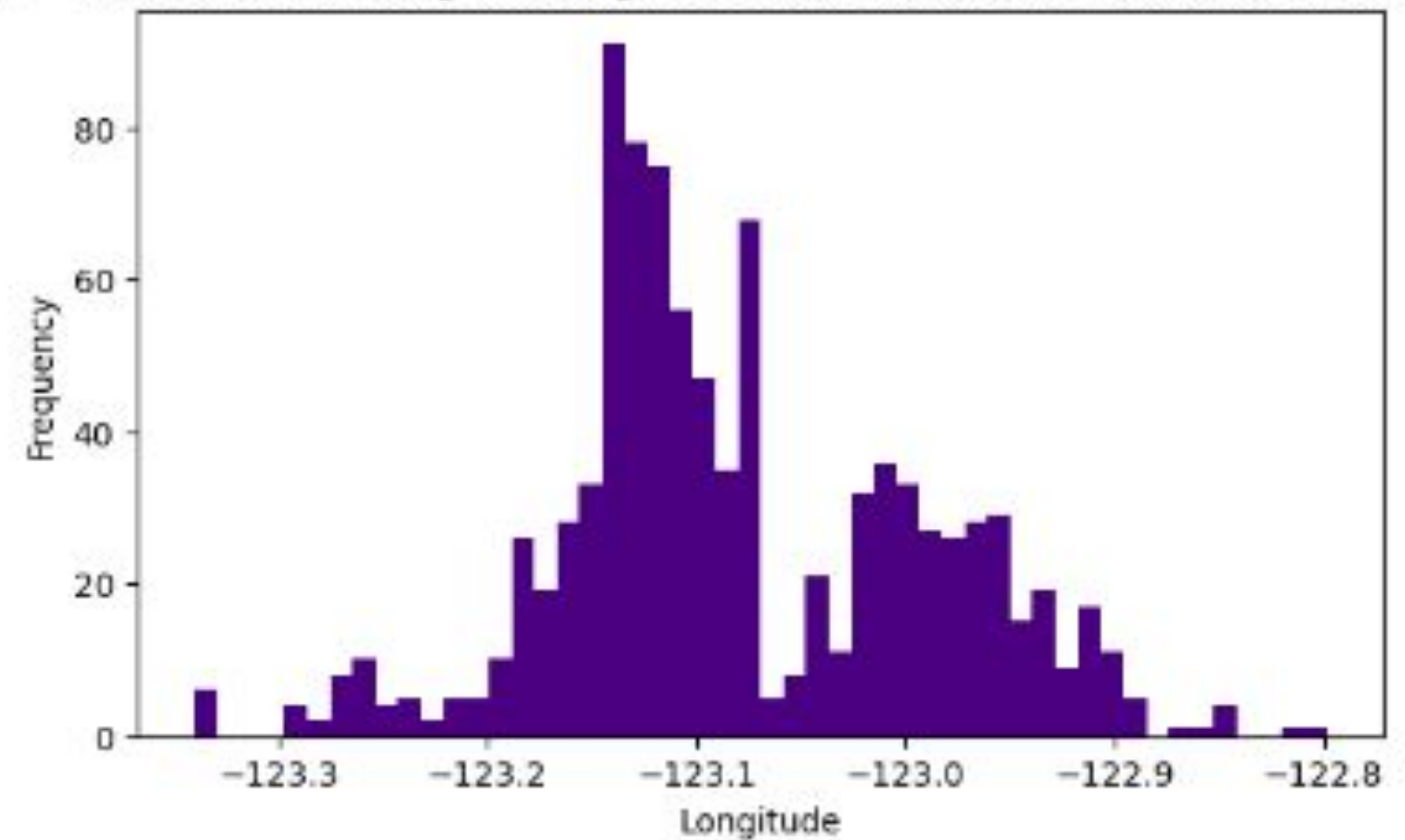
Dataframe Concatonation and PCA Transformation

LATITUDE AND LONGITUDE

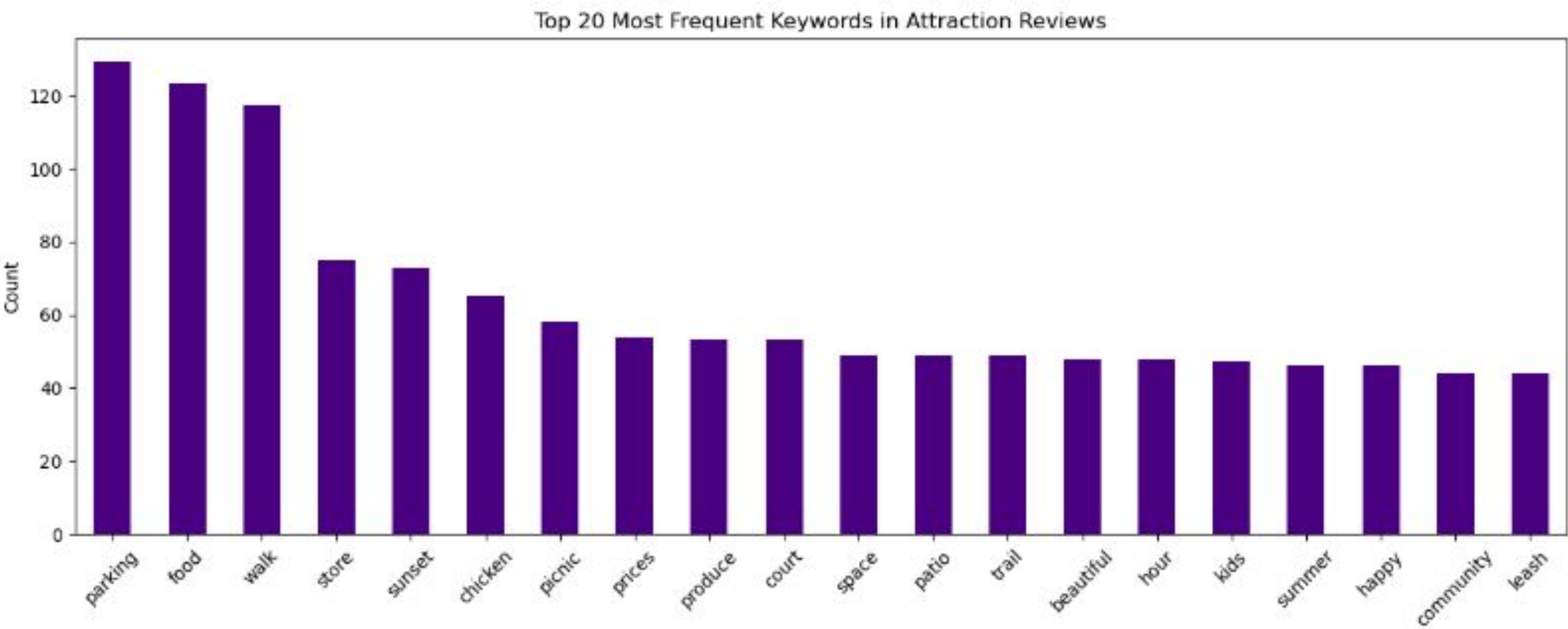
Distribution of Log Transformed Latitude for Attractions WITHOUT OUTLIERS



Distribution of Original Longitude for Attractions WITHOUT OUTLIERS

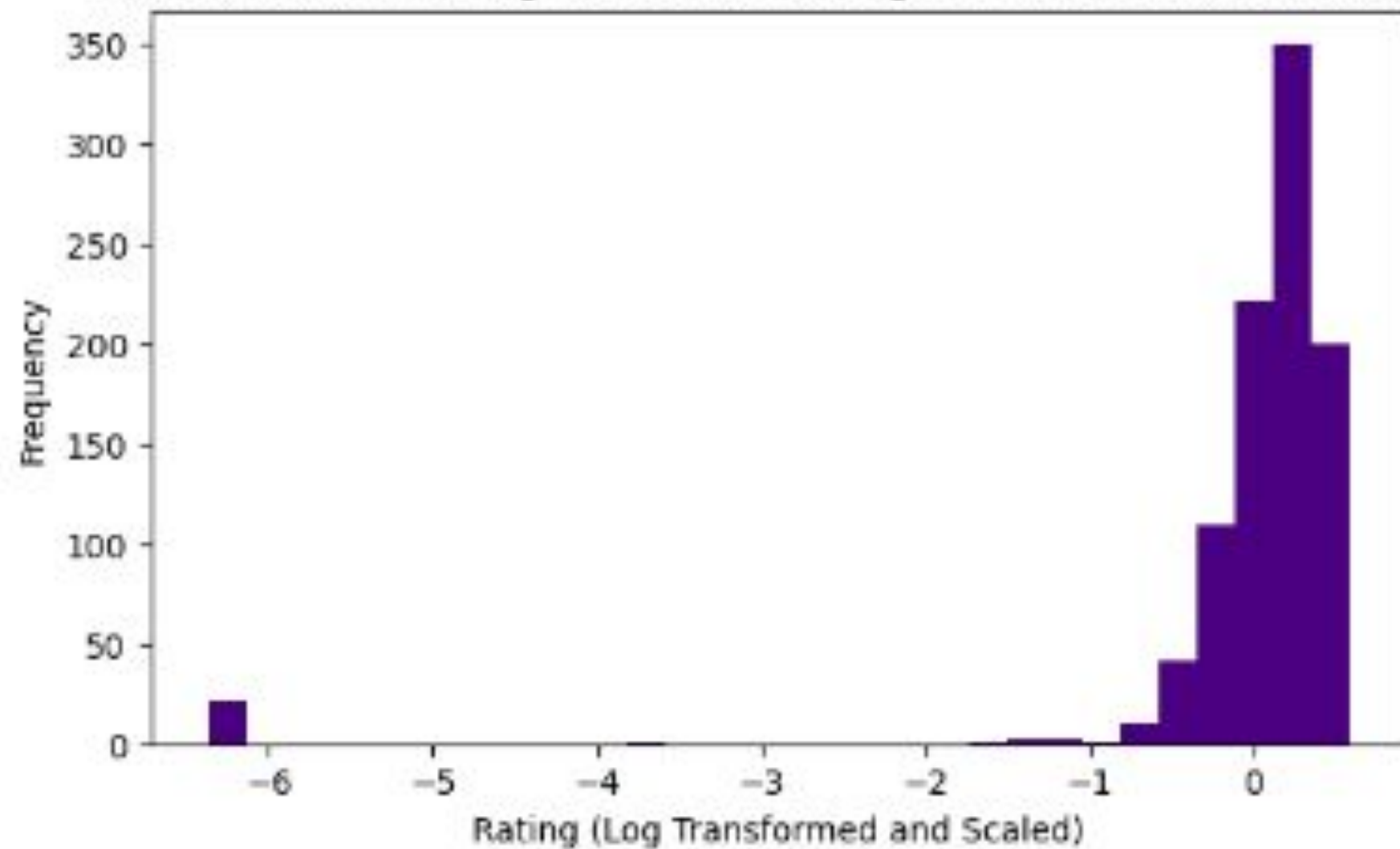


REVIEW KEYWORDS

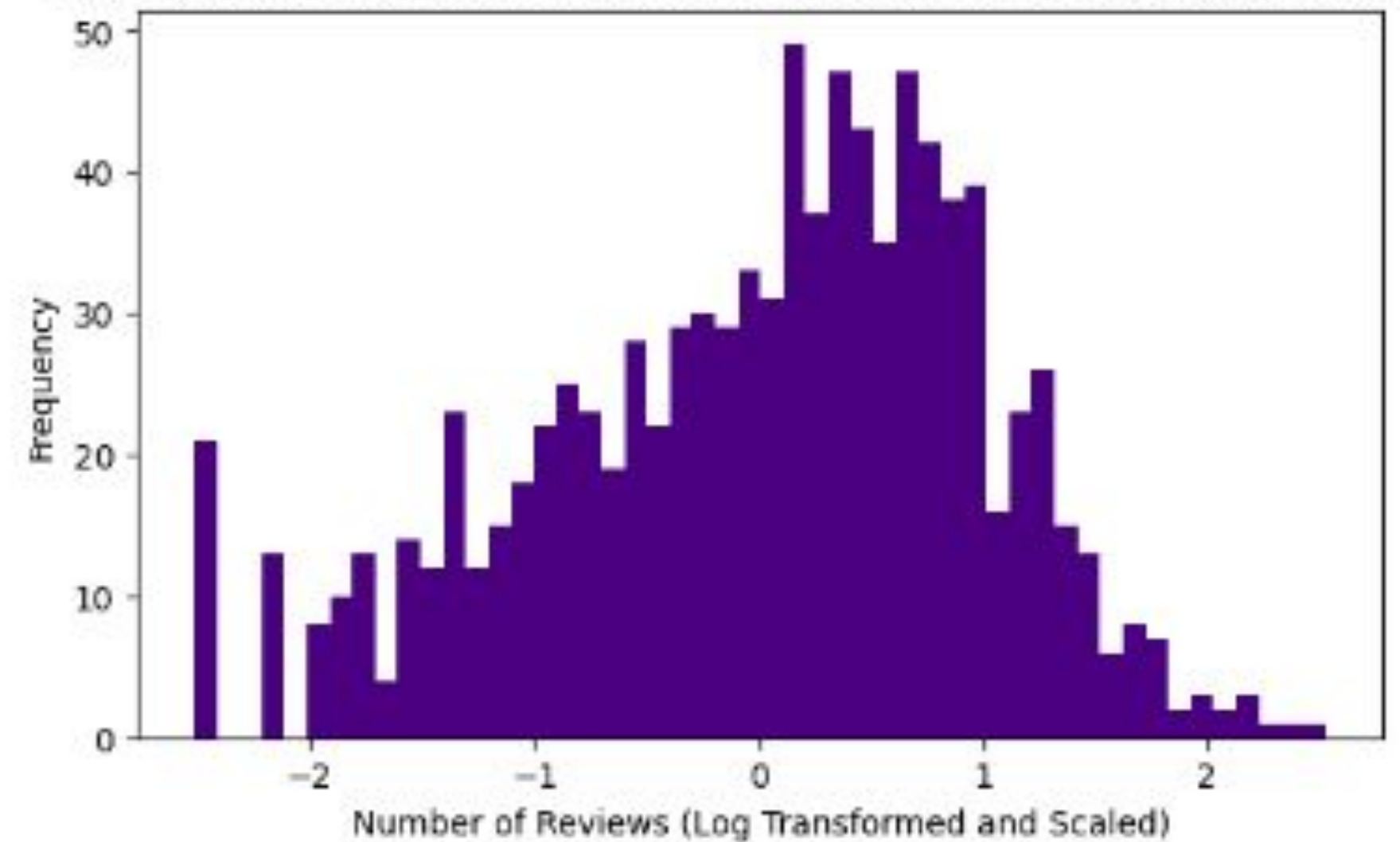


RATINGS AND REVIEWS

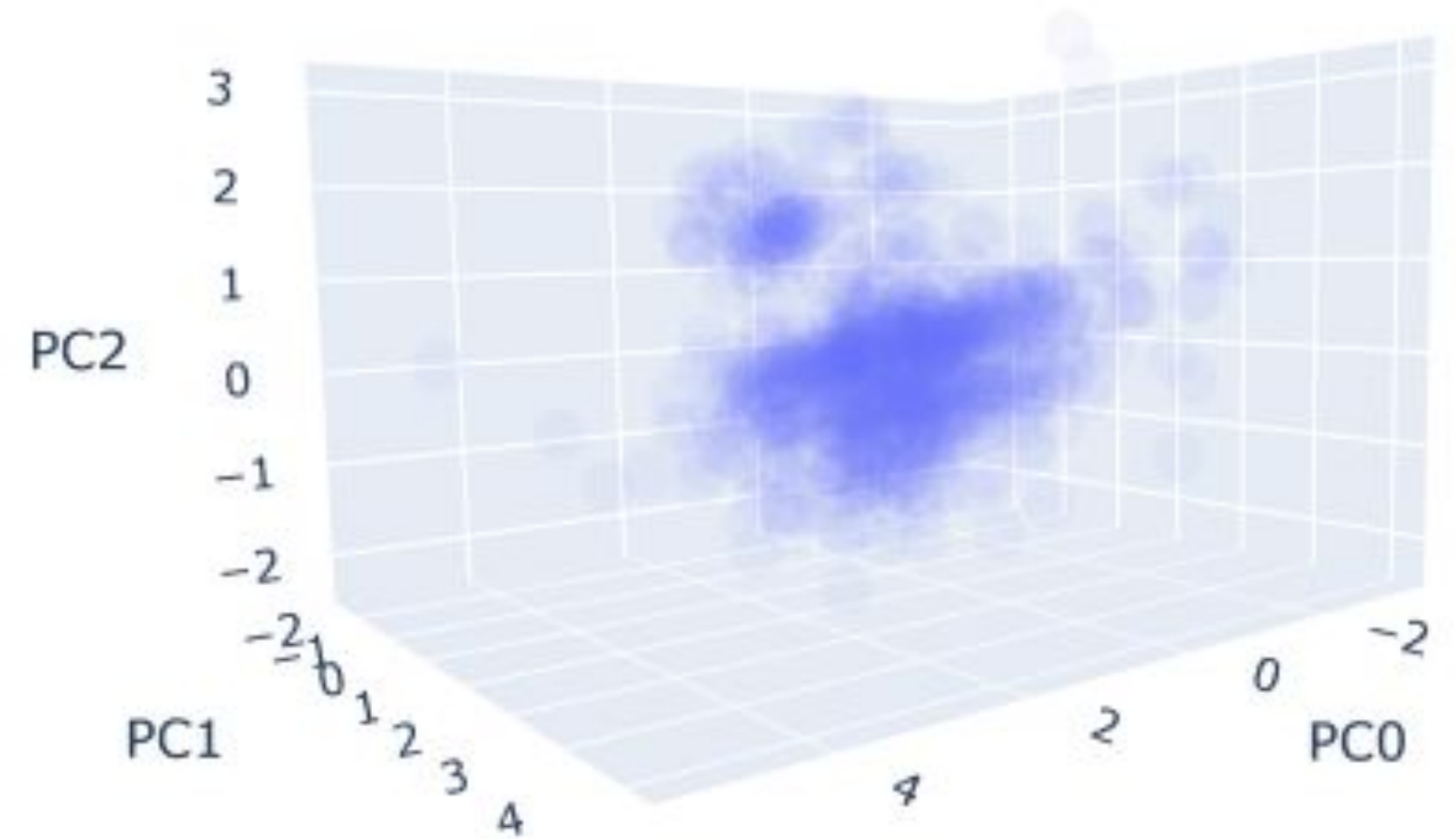
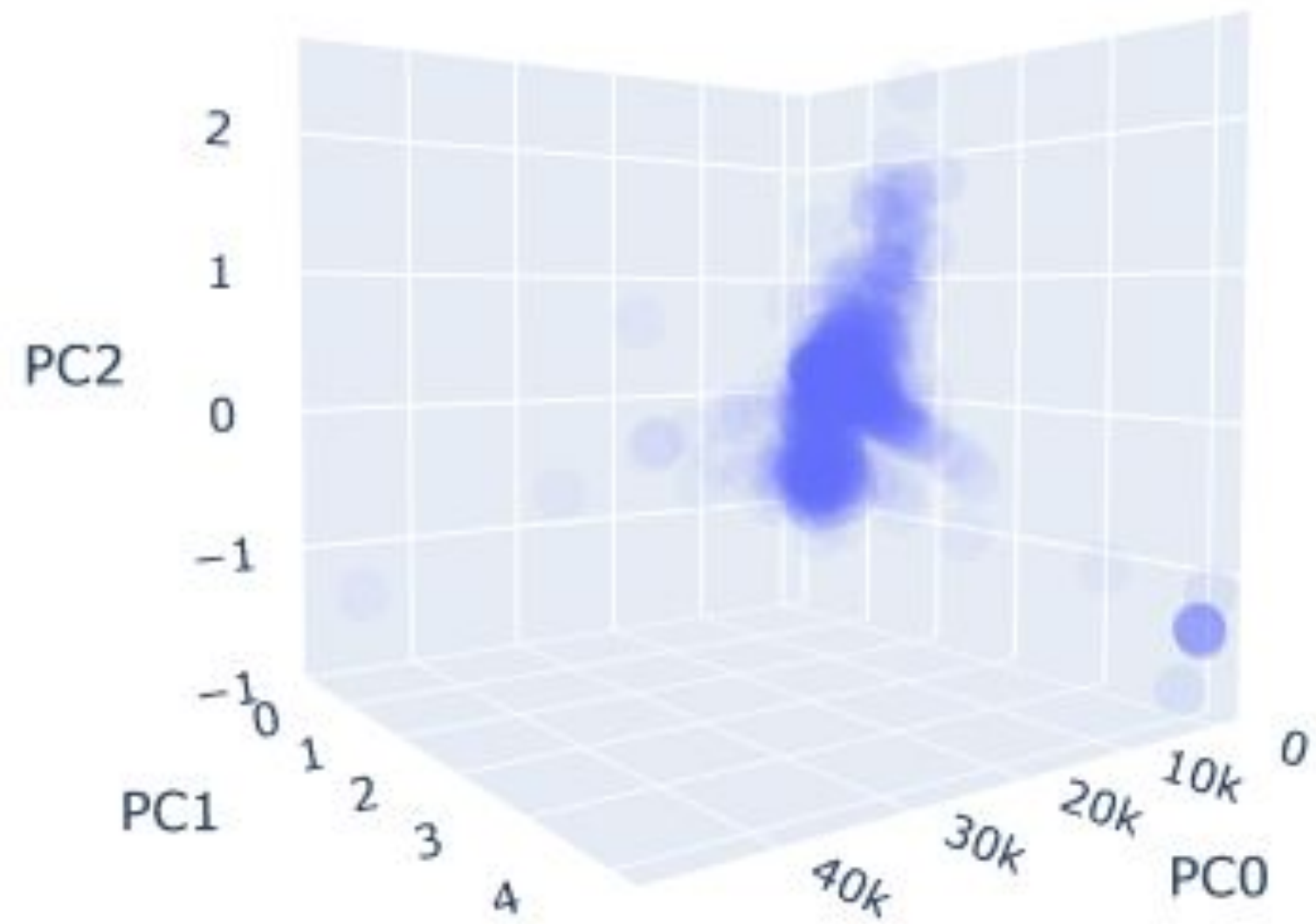
Distribution of Average Customer Ratings for Vancouver Attractions



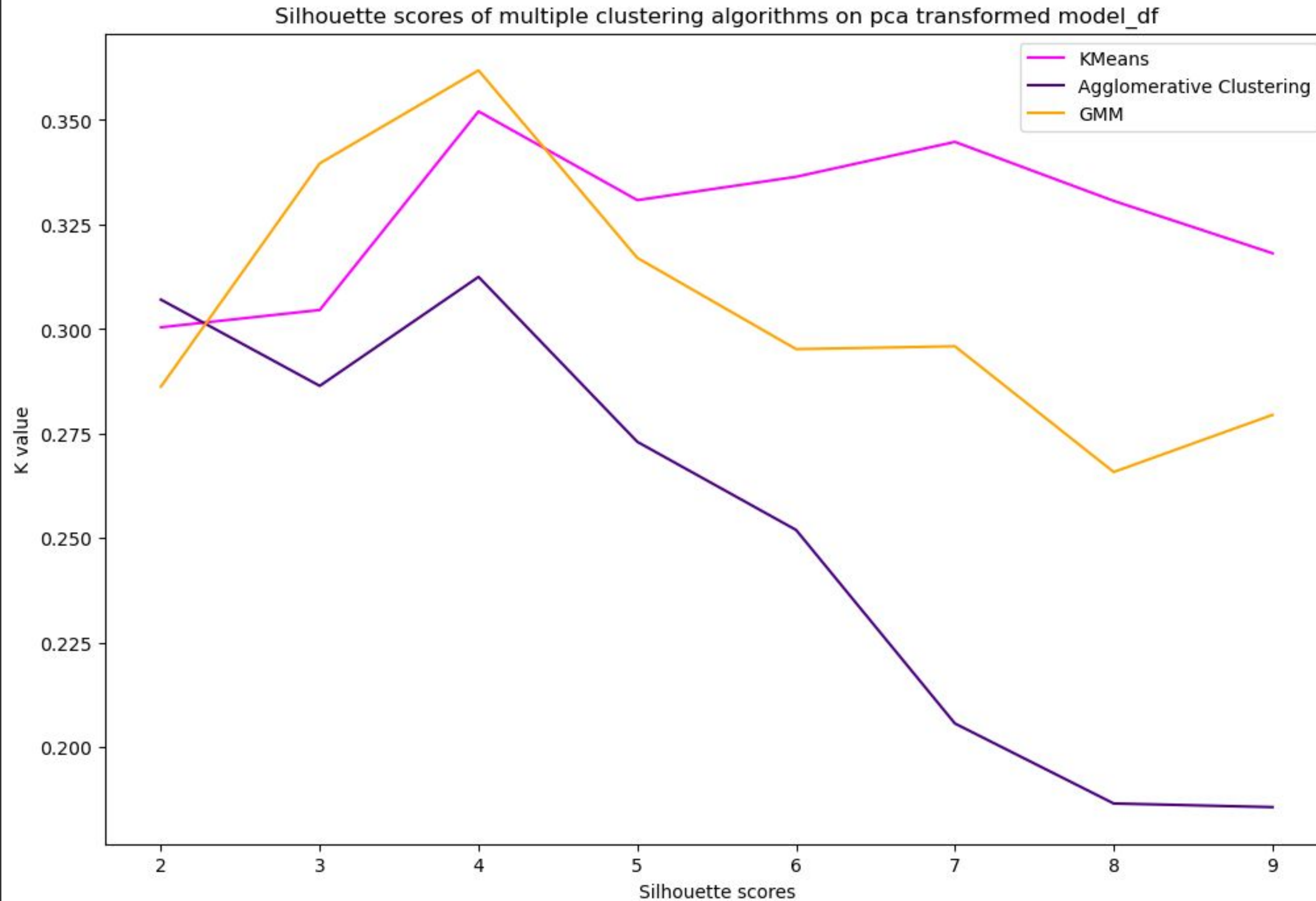
Distribution of Number of Customer Reviews for Vancouver Attractions



FINAL DATA MODEL



BASELINE MODEL



**Best Performance:
Gaussian Mixture Model**

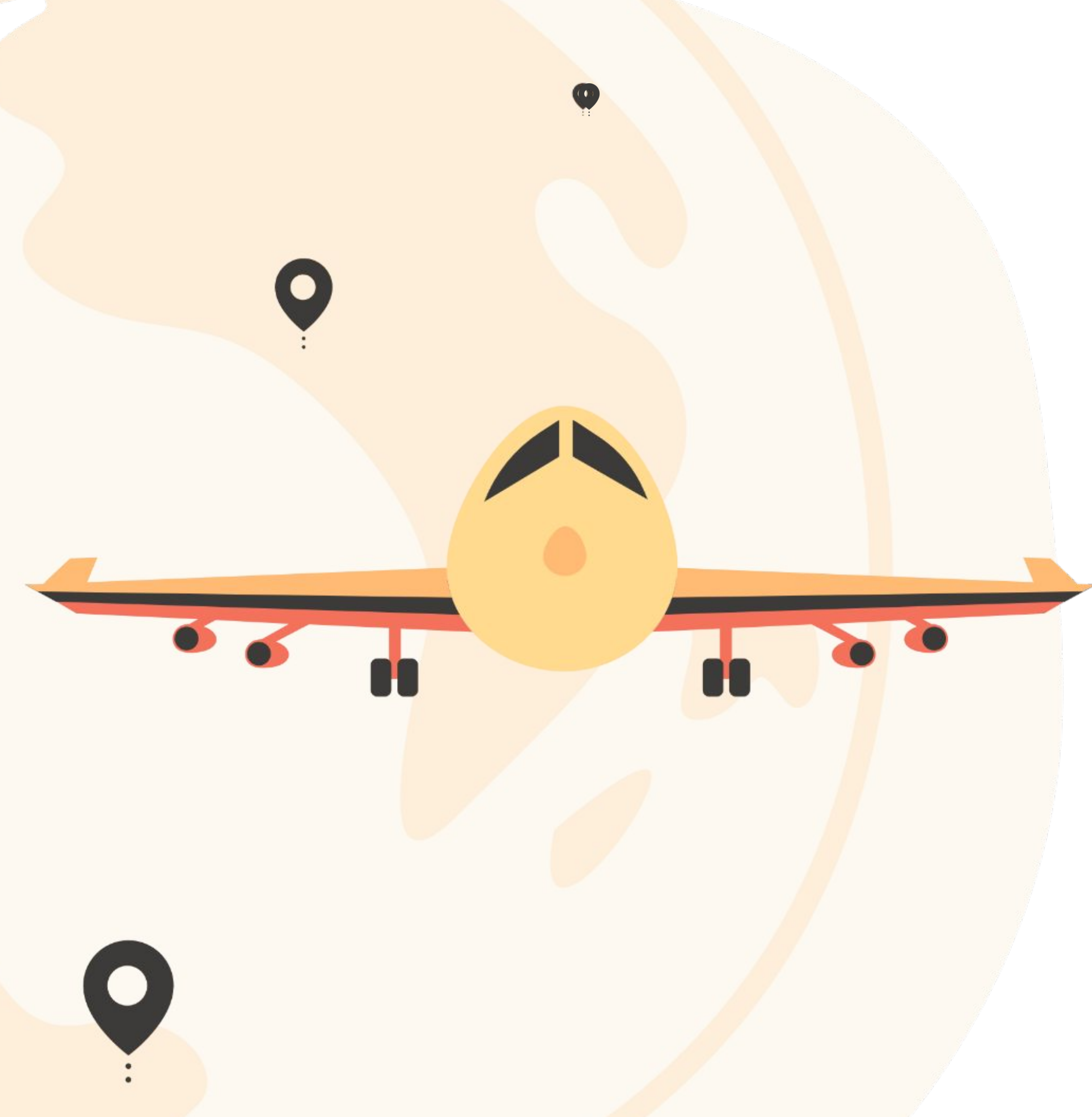
Number of clusters: 4

Score: ~0.365



NEXT STEPS

- **Feature Engineering**
 - Review Keywords
 - TF-IDF
 - Word2Vec
 - Rating
 - MaxAbsScaler
 - MinMaxScaler
 - Normalizer
- **Hyperparameter Tuning**
 - GaussianMixture Model
- **Content-Based Filtering**
- **Scheduling**



QUESTIONS?



CREDITS



This presentation template is free for everyone to use thanks to the following:

SlidesCarnival for the presentation template

Pexels for the photos

Happy designing!