Capstone 2 - Milestone

Yelp Recommendation System
By: Daniel Weissberger
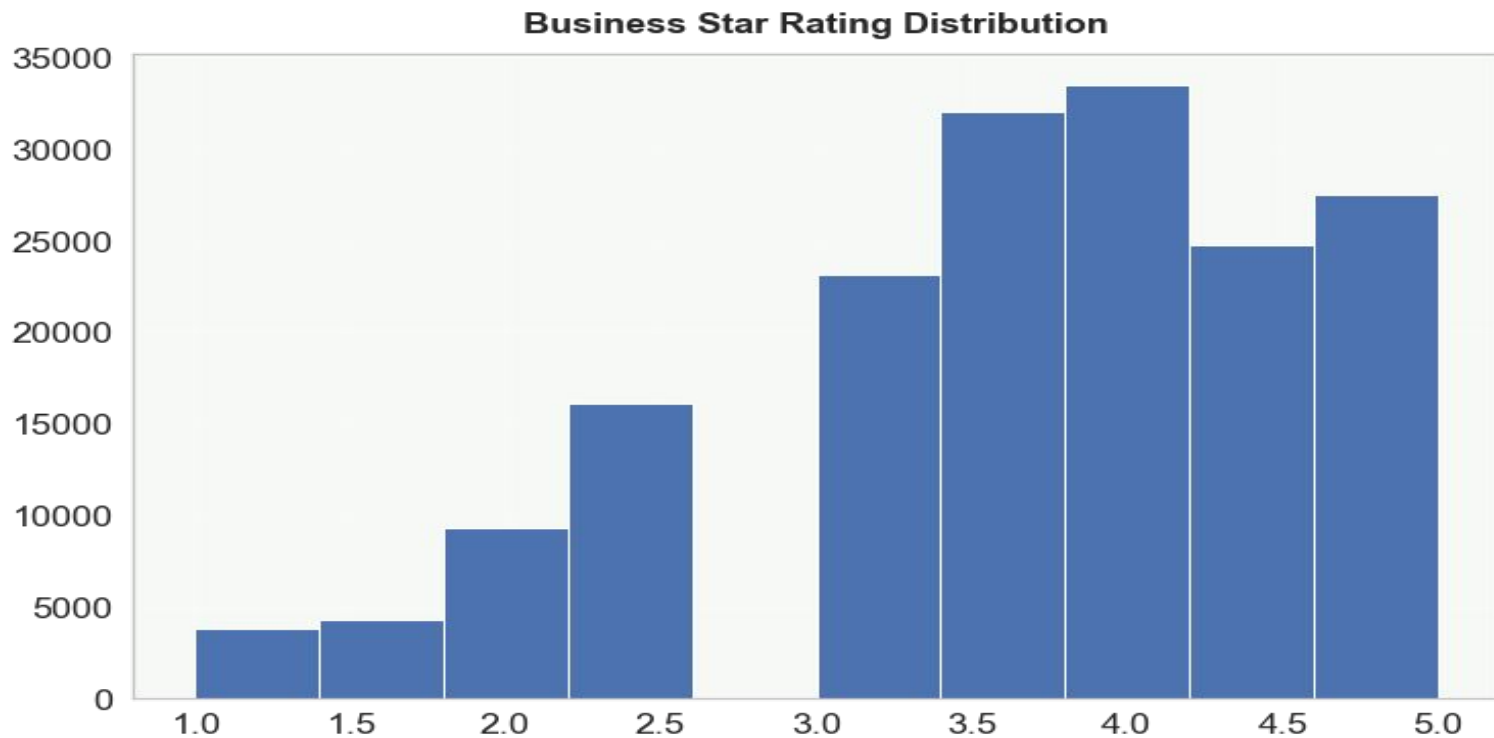
Data Wrangling

The following data files were provided in csv format:

File Name	File Size
yelp_business.csv	31,017 KB
yelp_business_attributes.csv	40,408 KB
yelp_business_hours.csv	13,542 KB
yelp_checkin.csv	4,935 KB
yelp_review.csv	3.7 GB
yelp_tip.csv	144,614 KB
yelp_user.csv	1.3 GB

EDA

Let's take a look at the star rating distribution:

Business Star Rating Distribution

Review Count Distribution (EDA continued)

Circos Plot in Kent, Ohio

Top 10 Active Users Found
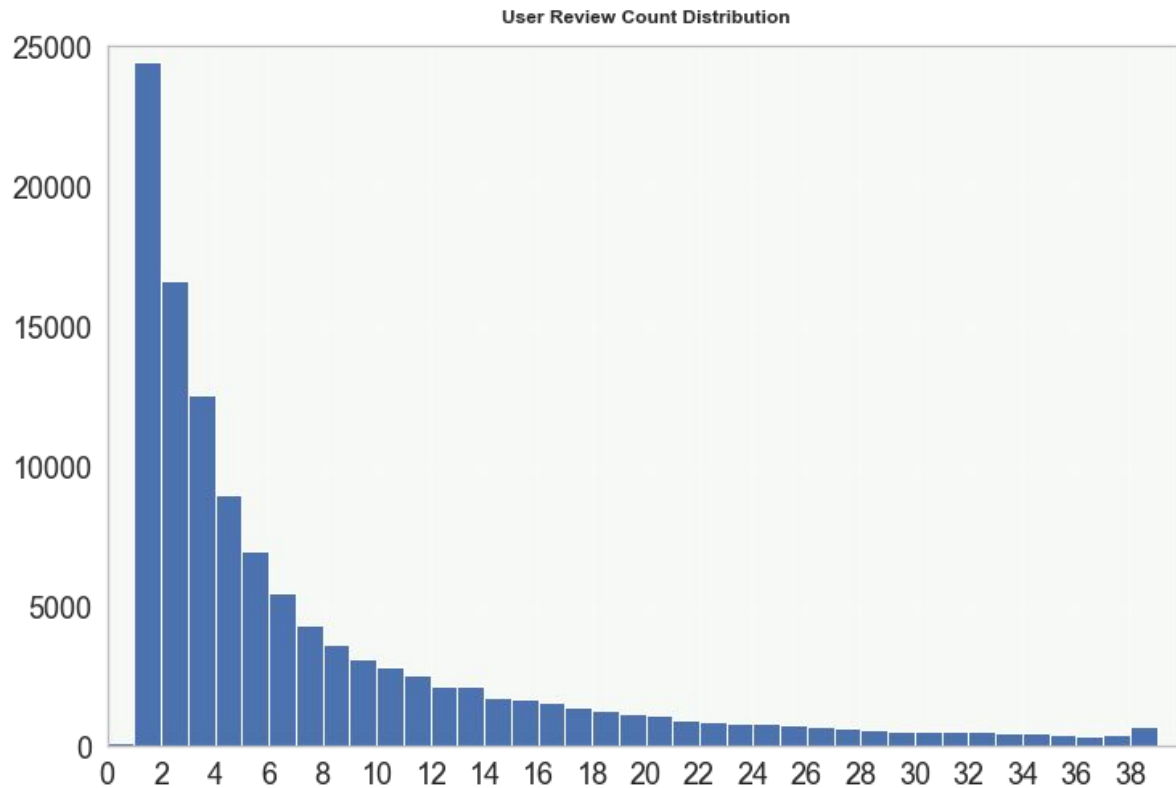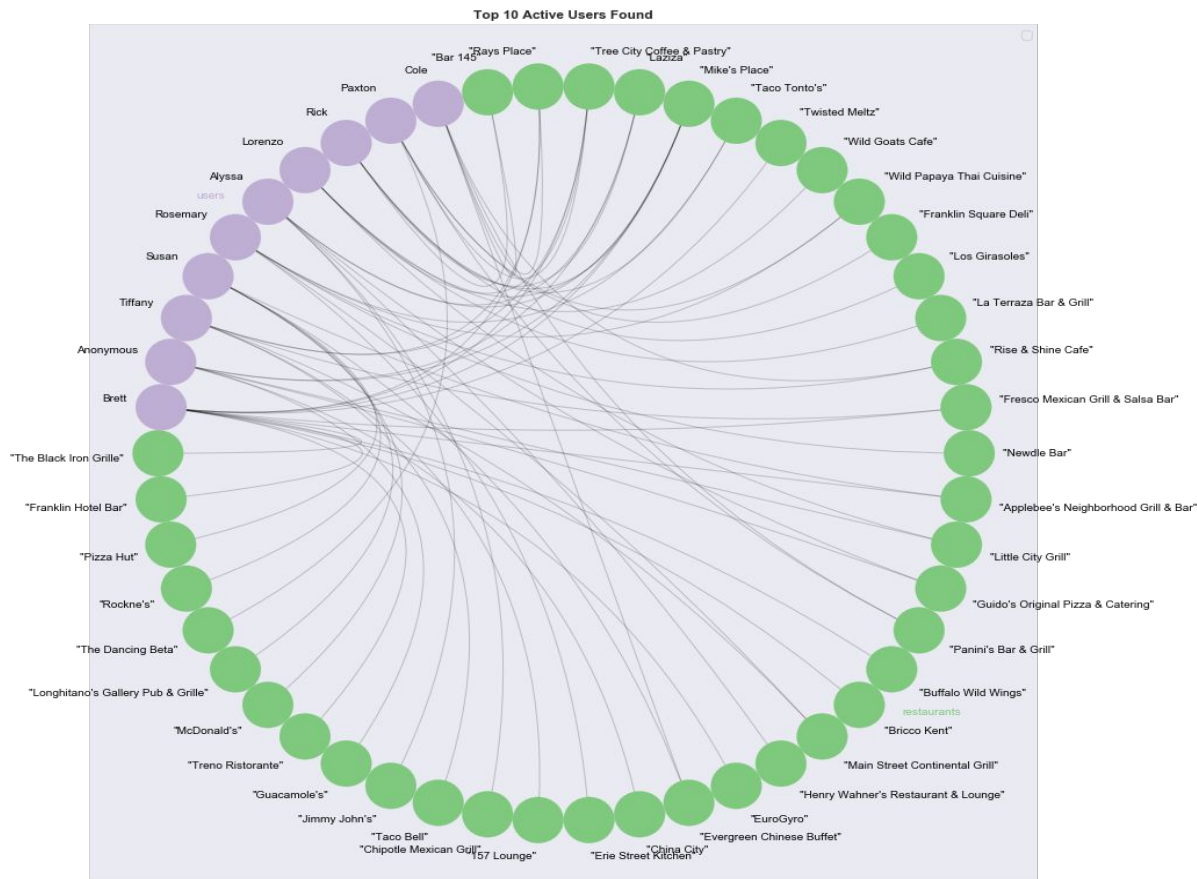
Recommendation System - Content Filtering Results

Accuracy	.526
Precision	.742
Recall	.546

Matrix Factorization - Explained

Rating Matrix

	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

=

User Matrix

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

X

Item Matrix

W	X	Y	Z
1.5	1.2	1.0	0.8
1.7	0.6	1.1	0.4

Matrix Factorization - Results

Accuracy	.719
Precision	.735
Recall	.952

Accuracy by Iteration