

Business Analysis

on



- NAIQING CAI
- JITIAN ZHAO
- ZIHAO LI
- YAOBIN LING

STAT 628 Module 3

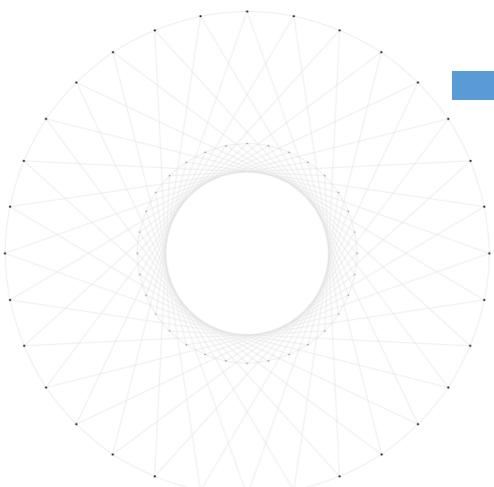
PRESENTED BY GROUP 2



CONTENTS

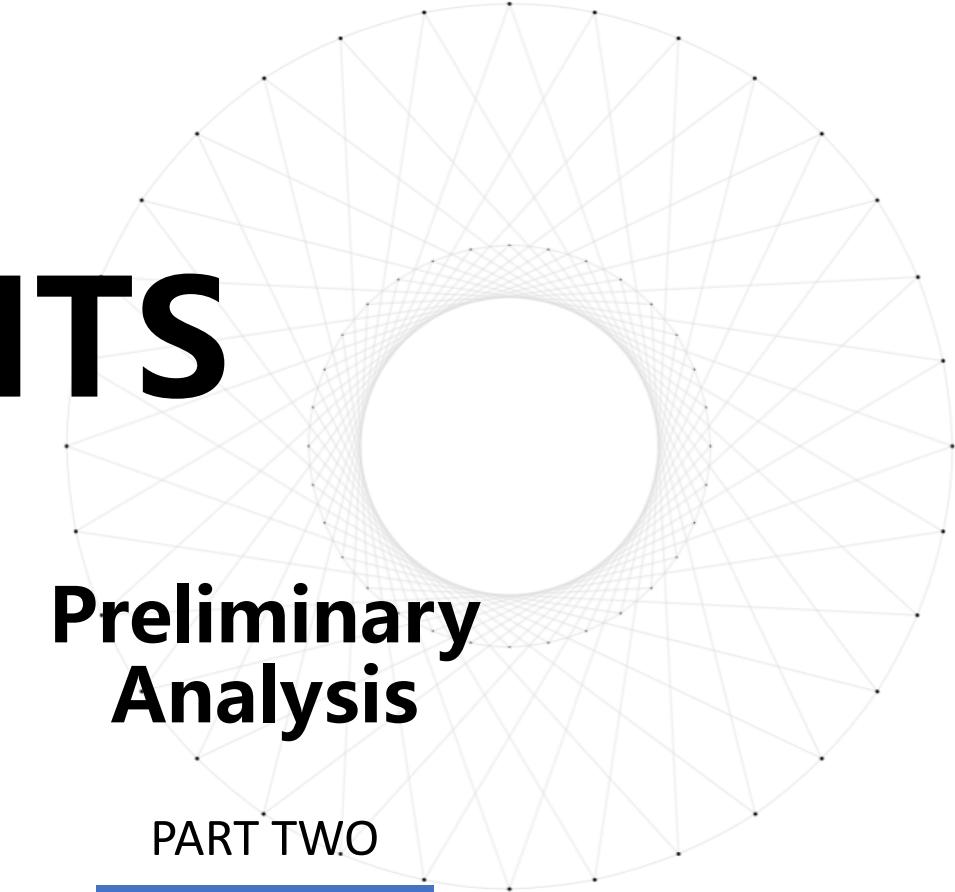
Data Analysis Plan

PART ONE



Preliminary Analysis

PART TWO



Data Analysis Plan

PART ONE



PART ONE Data Analysis Plan

Background

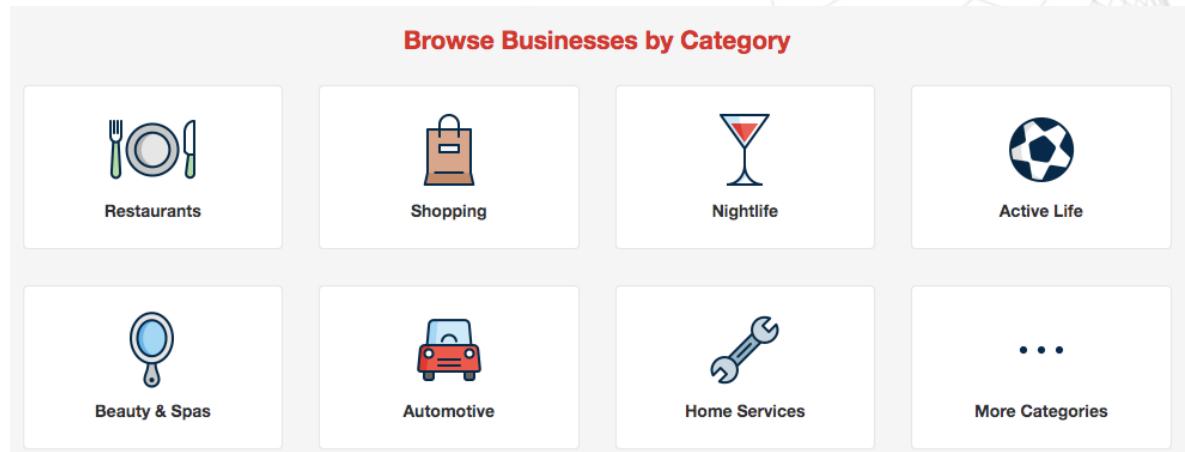
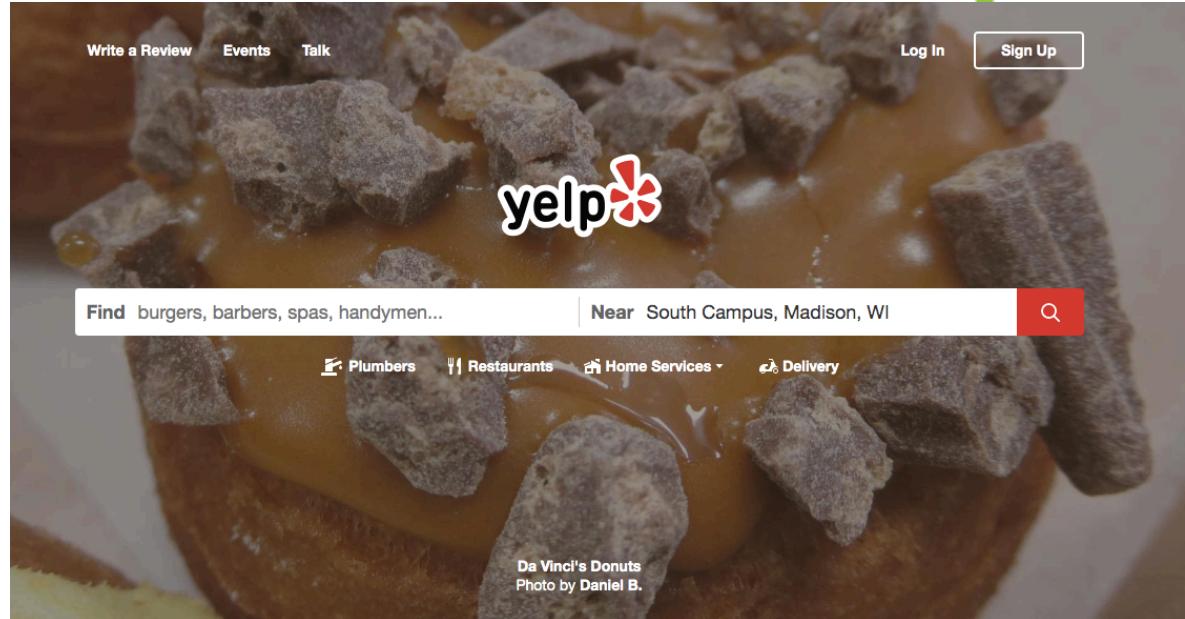
Information in Data



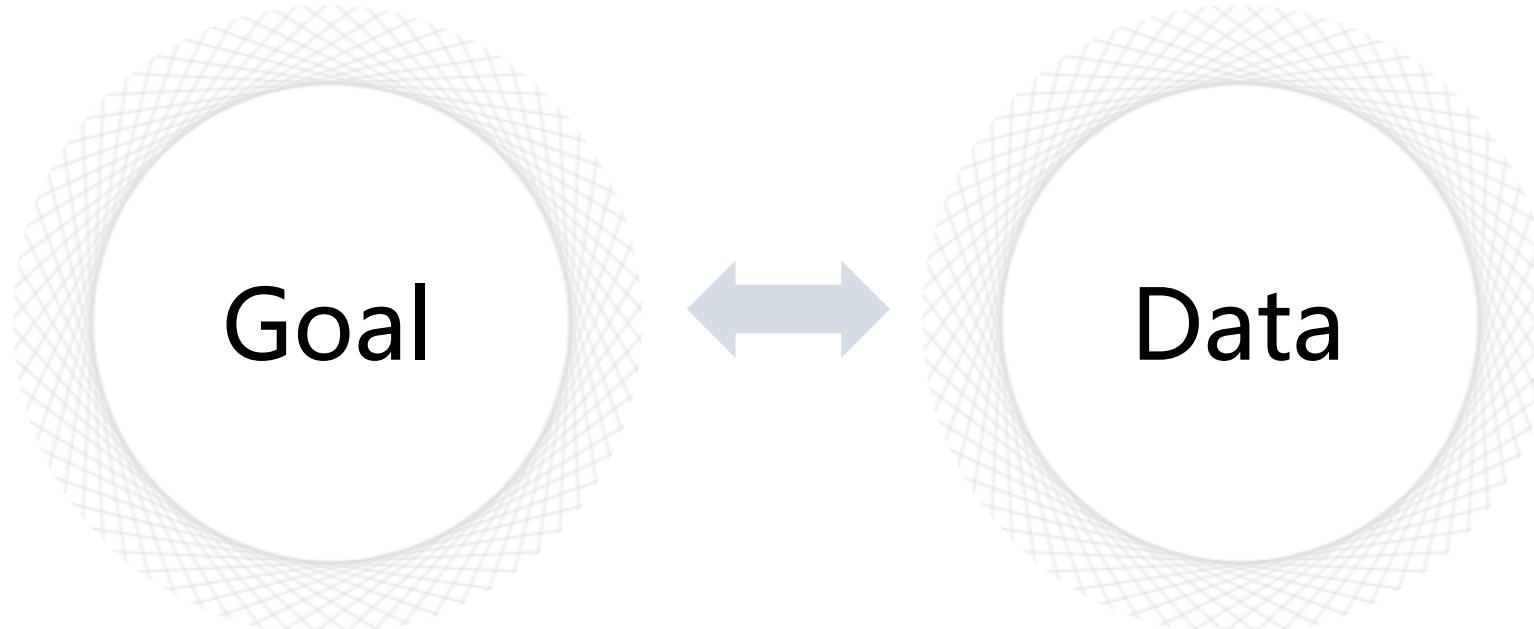
Statistical Analysis



Data-driven and Actionable Decisions

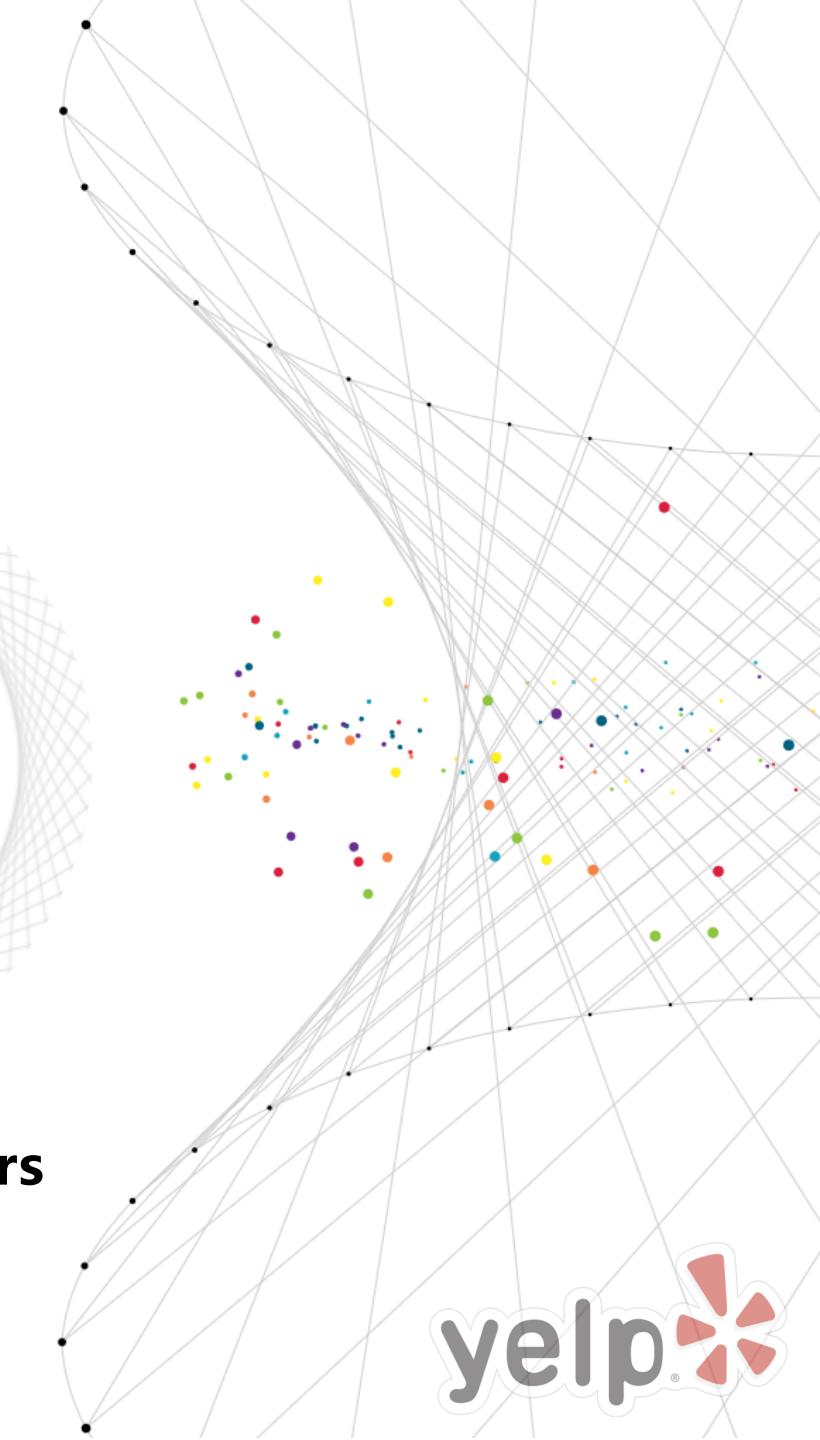


Identify Business Problems



Data-driven and Actionable Decisions

Restaurant——Bars



PART ONE Data Analysis Plan

Step 1 Data Preprocessing

Data Overview

Text Cleaning

Data Visualization

Step 3 Learning Algorithm

Topic Modeling—LDA

Step 5 Application

Shiny App

Decision Making

Step 2 Feature Extraction

TF-IDF

Step 4 Model Evaluation

Performance Metrics

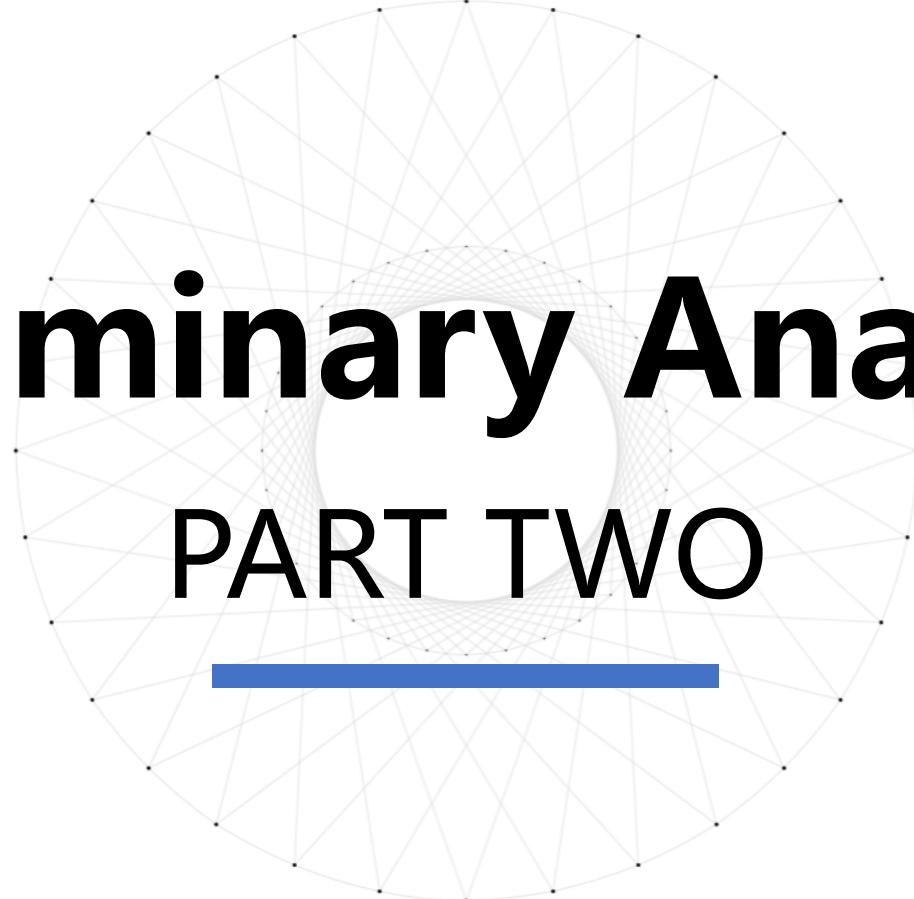
Diagnostics

WORK FLOW



Preliminary Analysis

PART TWO



Data Overview——Data Structure

Business

Business id	is_open	categories	Attributes	Address City & State Postal code Latitude & Longitude	Stars
-------------	---------	------------	------------	--	-------

Review

Review id	User id	Business id	stars	useful	funny	cool	text
-----------	---------	-------------	-------	--------	-------	------	------

User

User id	Review count	Start & Elite	useful	funny	cool	fans	friends
---------	--------------	---------------	--------	-------	------	------	---------

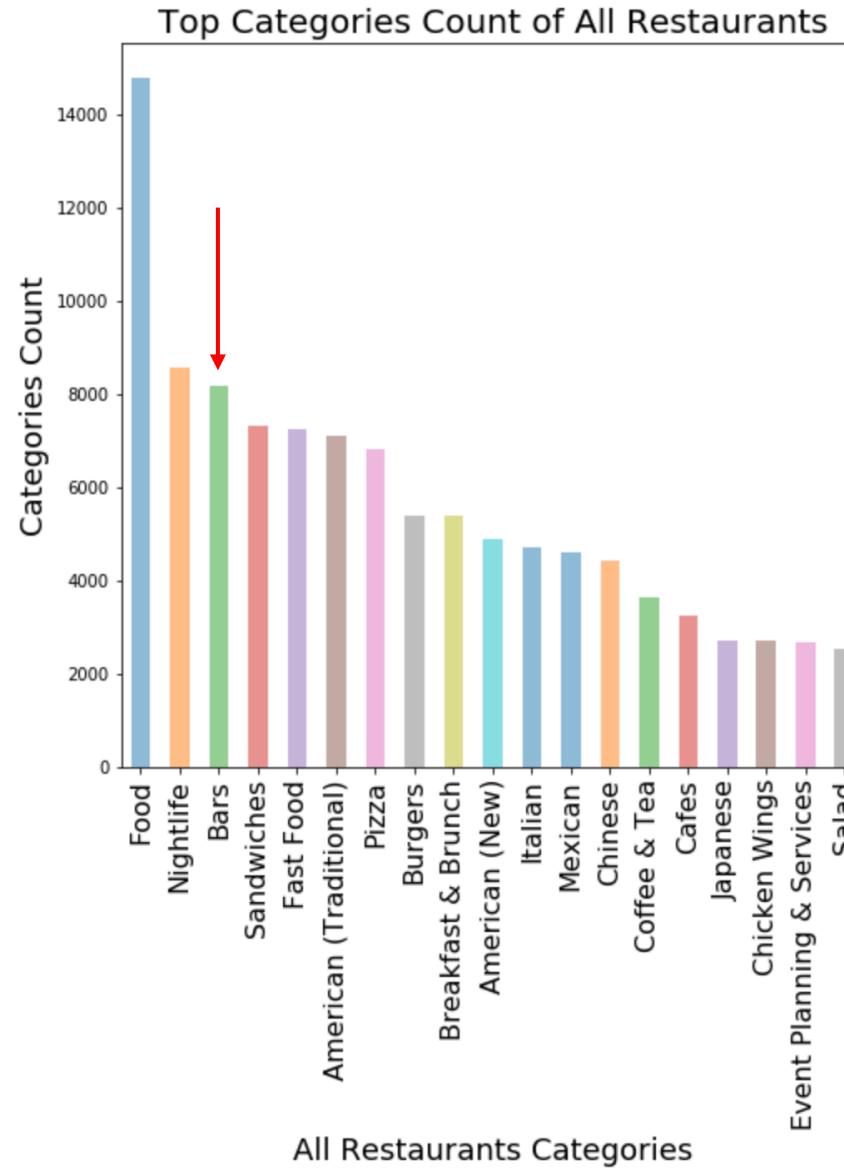
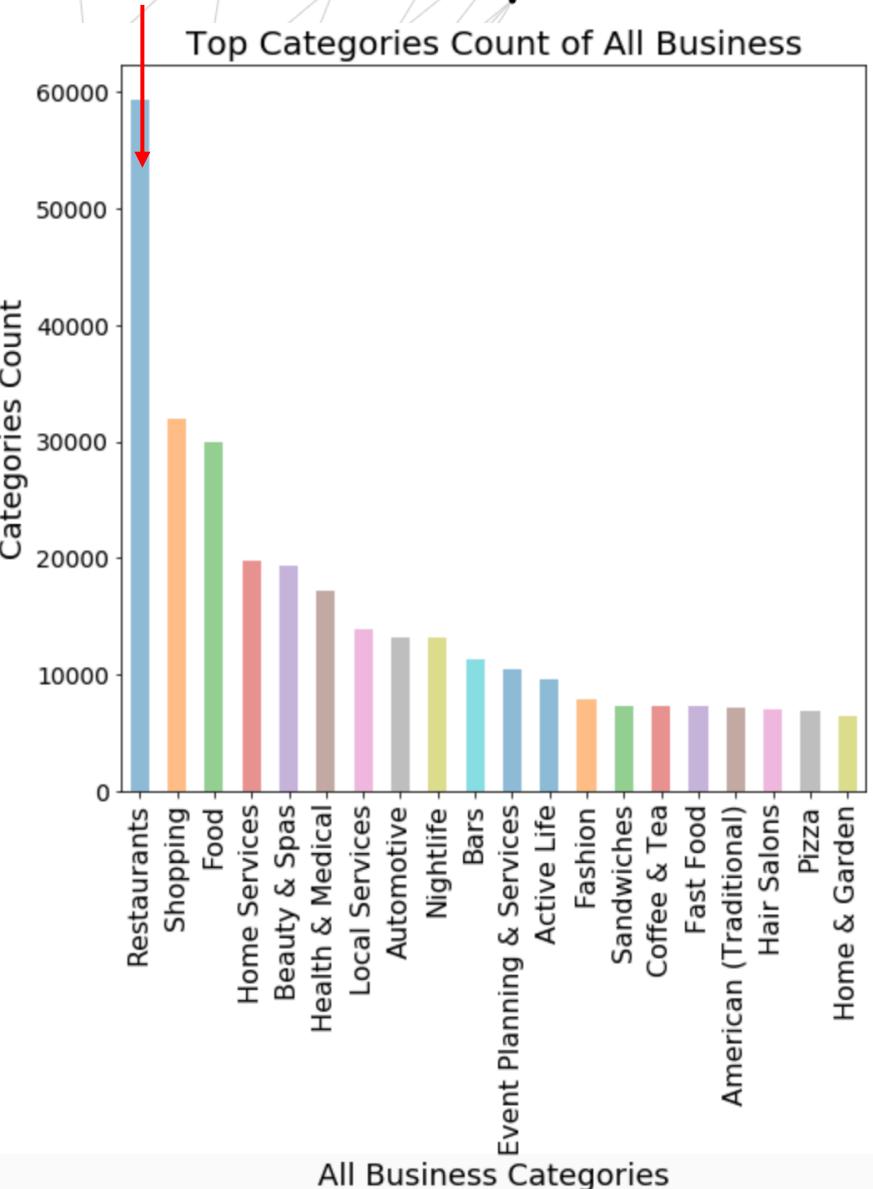
Tip

User id	Business id	text
---------	-------------	------



Data Overview—Categories

PART TWO Preliminary Analysis

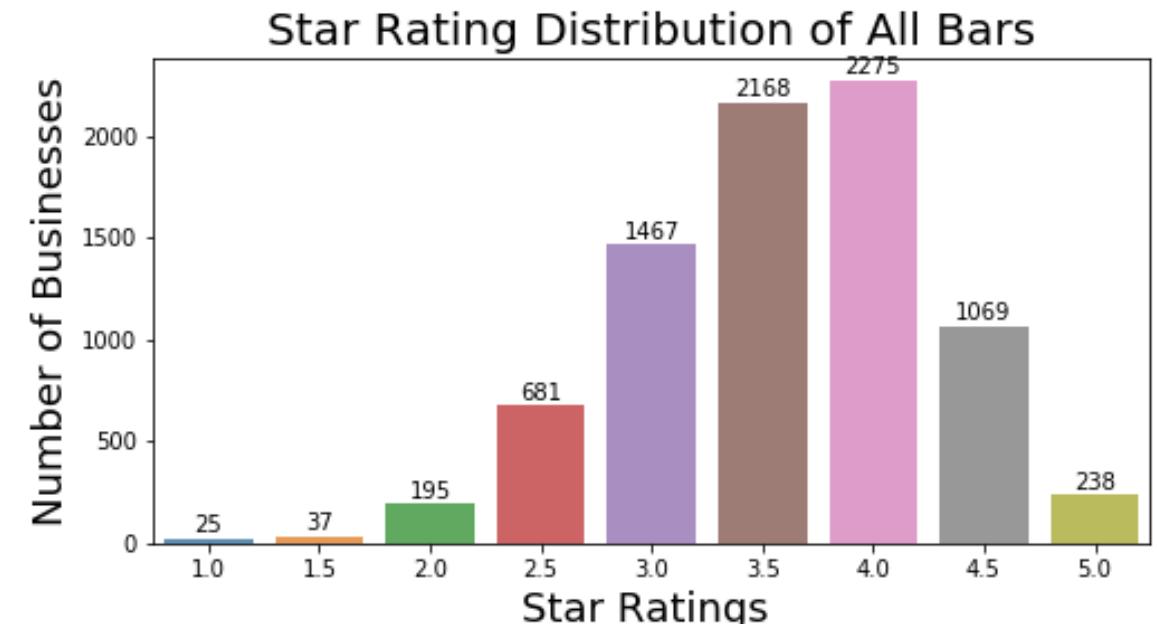
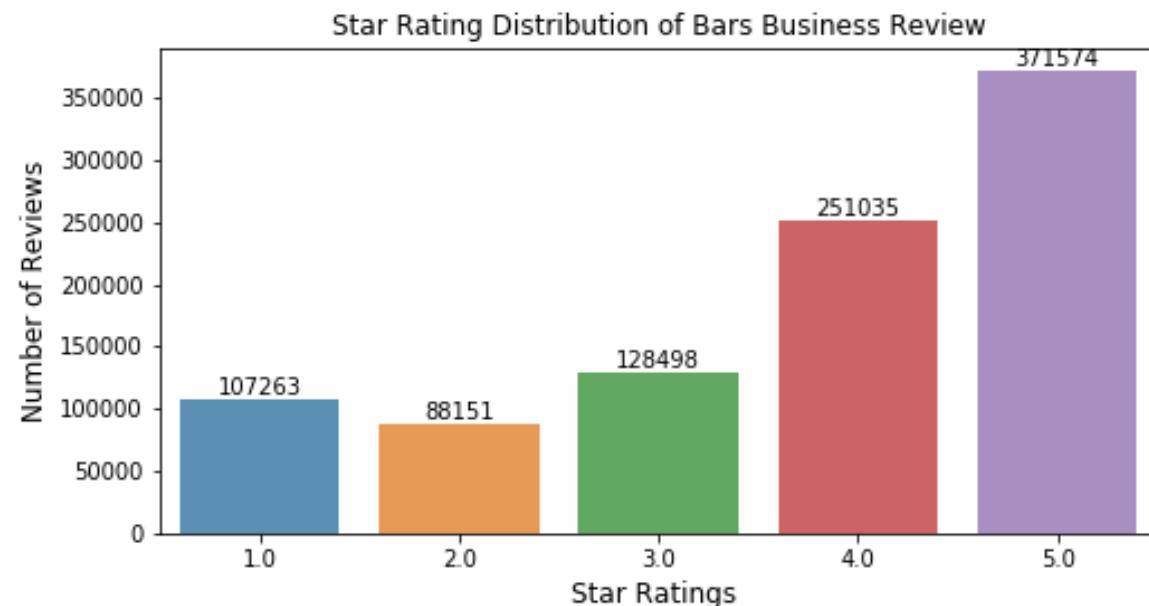


Data Overview—Cleaned-up Data

File	Original	Cleaned-up Bar Business
Business	192,609	8,155
Review	6,685,900	946,522
User	1,637,138	453,004
Tip	1,223,094	170,127



Data Overview—Star Rating



Text Cleaning

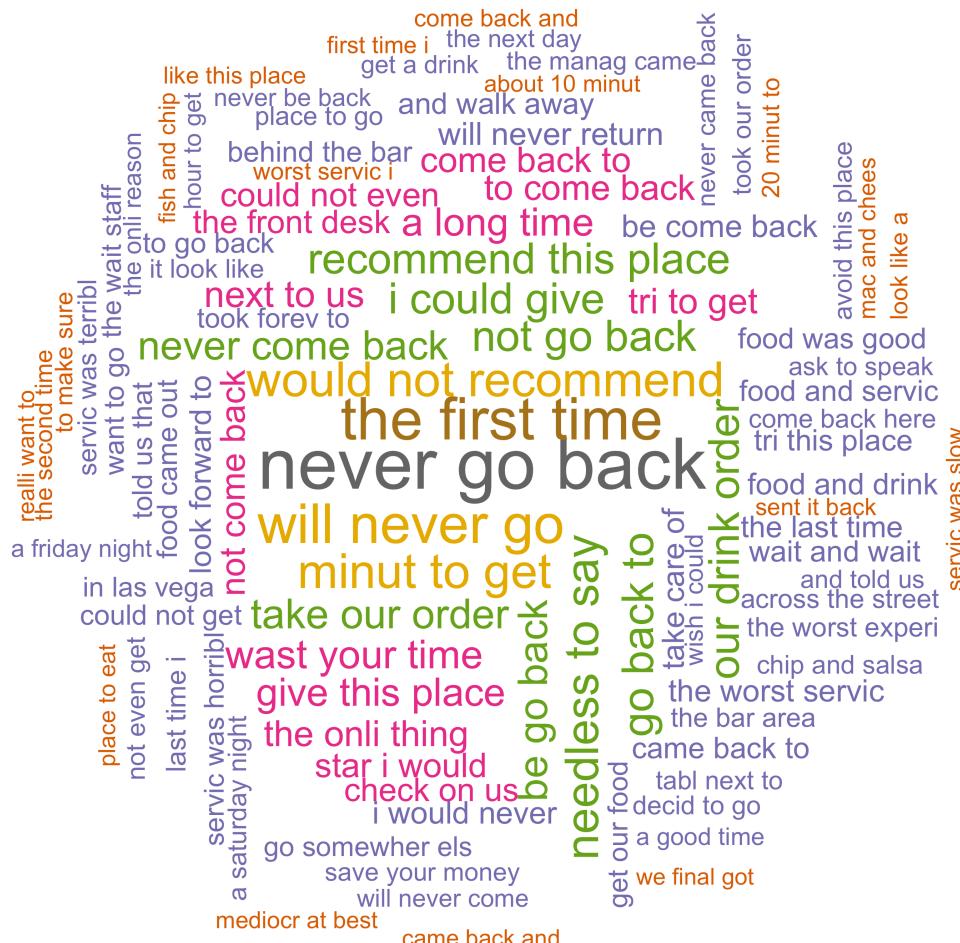
- Translation
- Negative Postfix Transformation
- Tokenization
- Removing Punctuation
- Removing Stop Words
- Converting Text to Lower Case
- Misspelling
- ...

Data Visualization—Word Cloud (3-gram)

5 Star Reviews

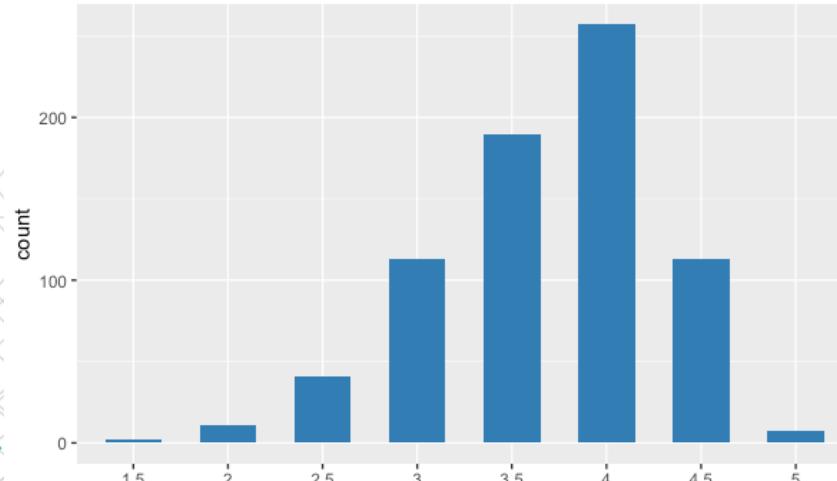


1 Star Reviews

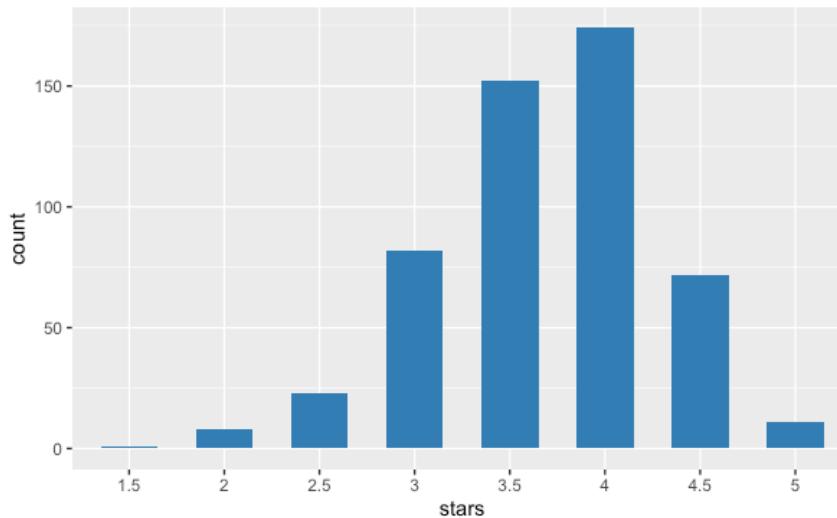


Data Visualization—Attributes (Music)

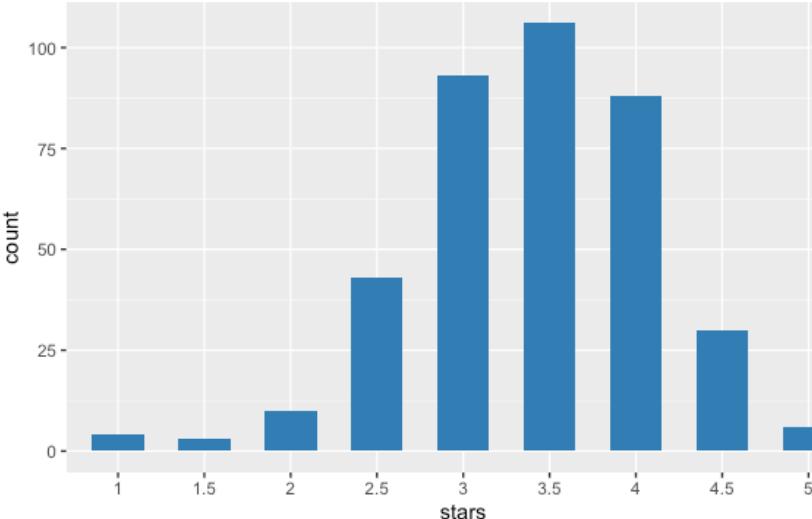
Background Music



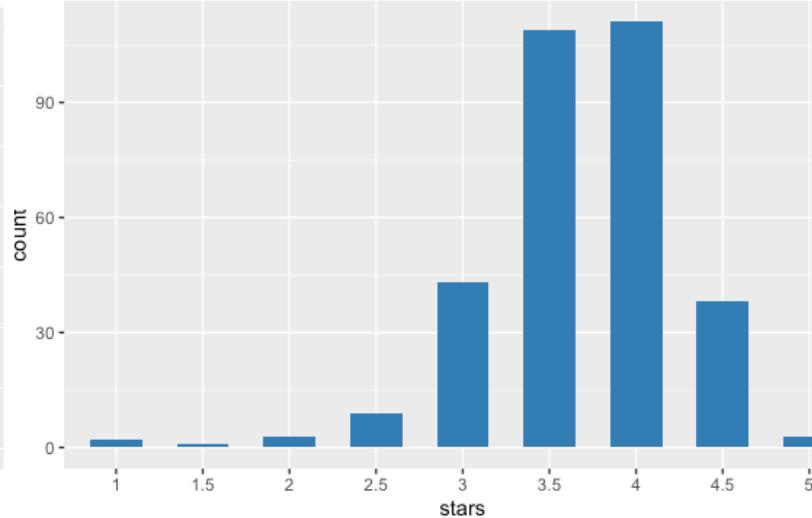
Live



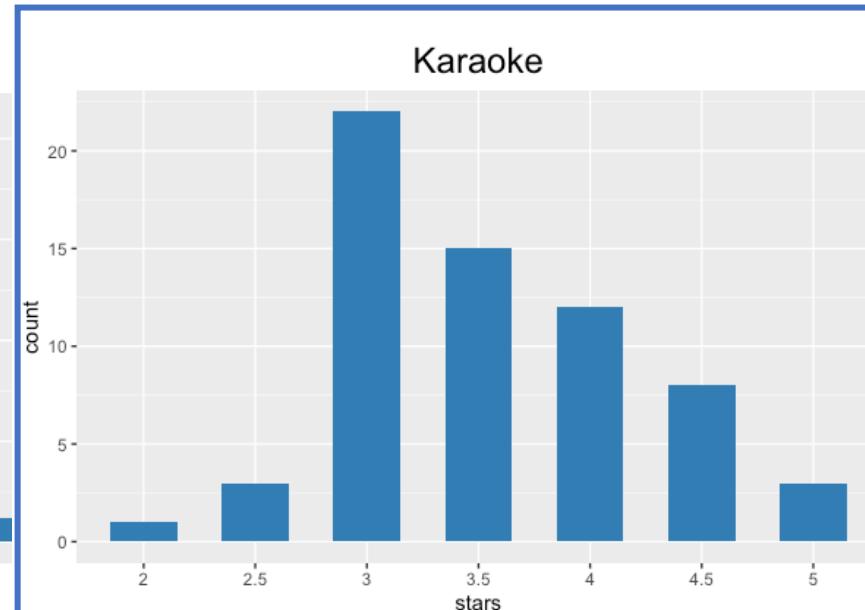
DJ



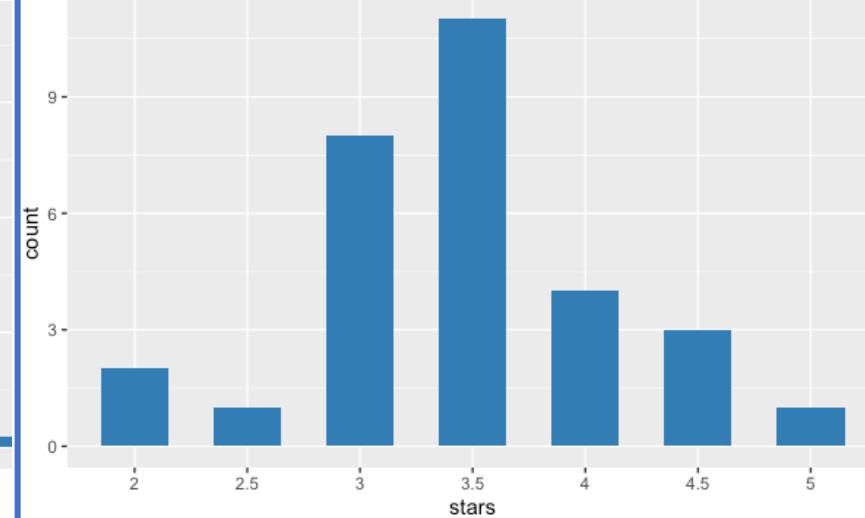
Jukebox



Karaoke

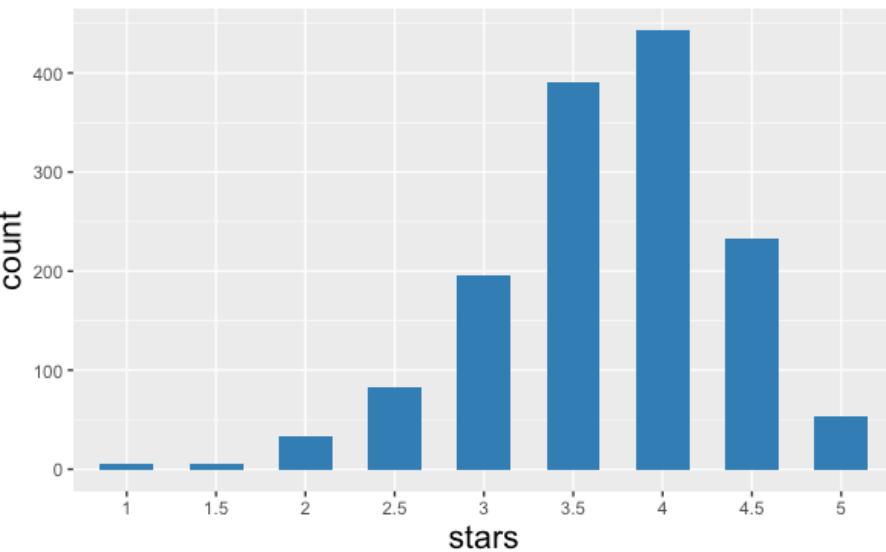


Video

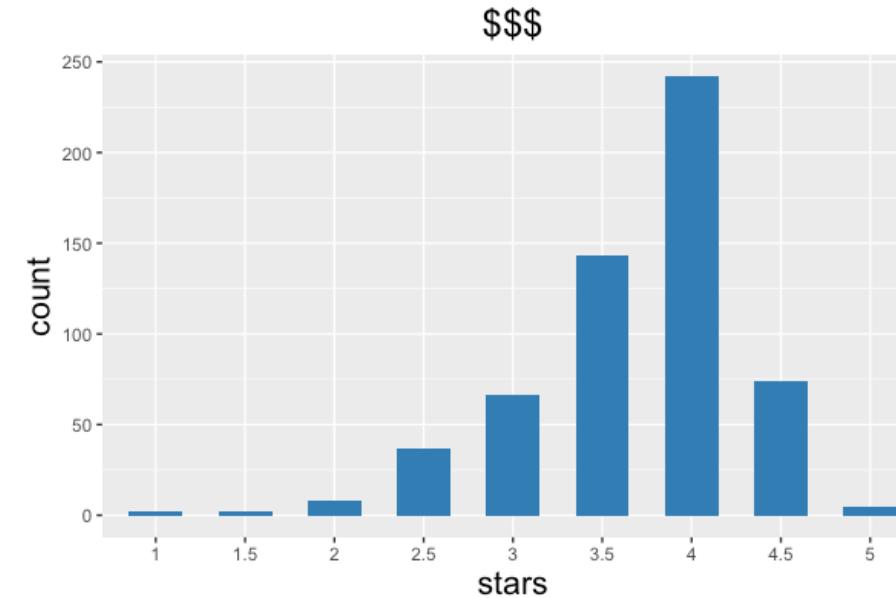


Data Visualization—Attributes (Price)

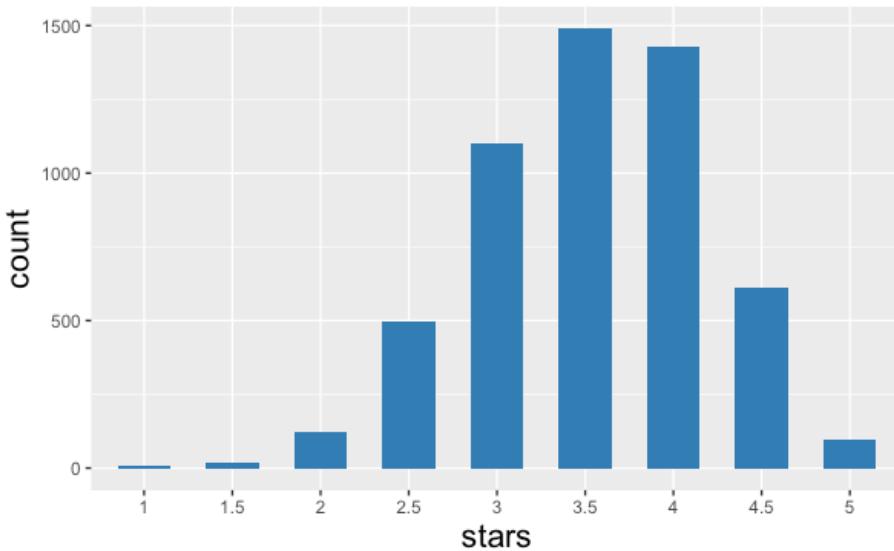
\$



\$\$\$

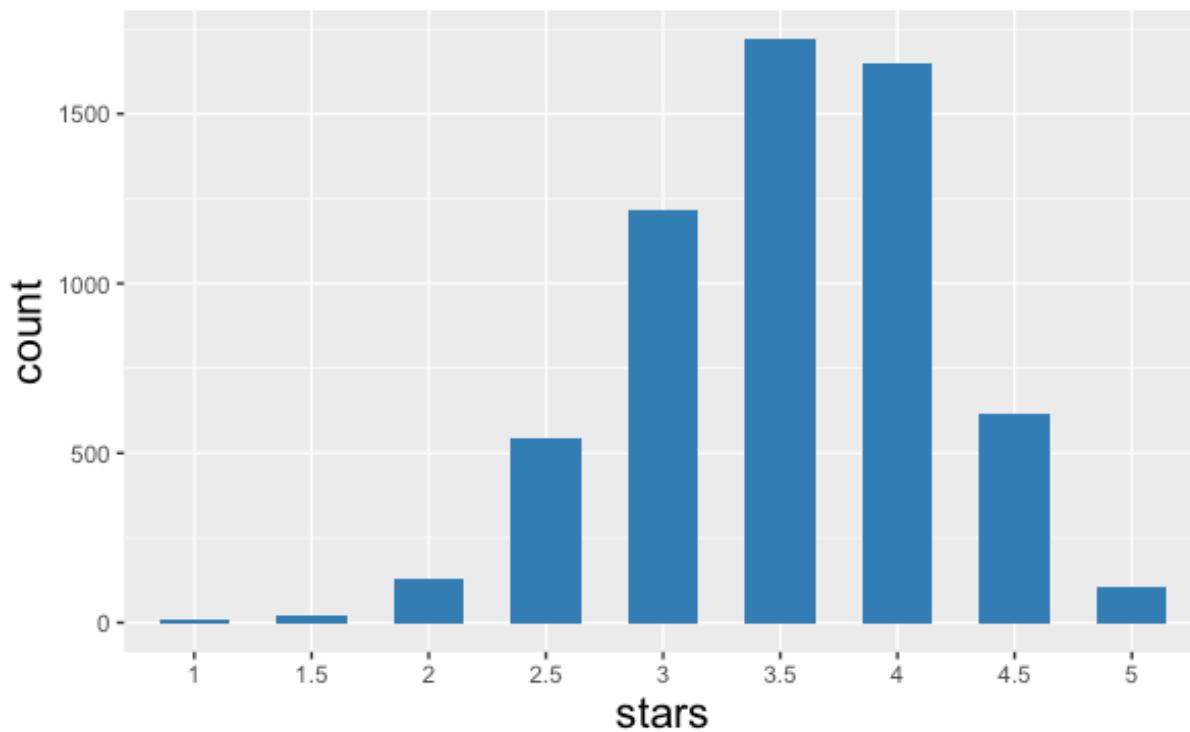


\$\$

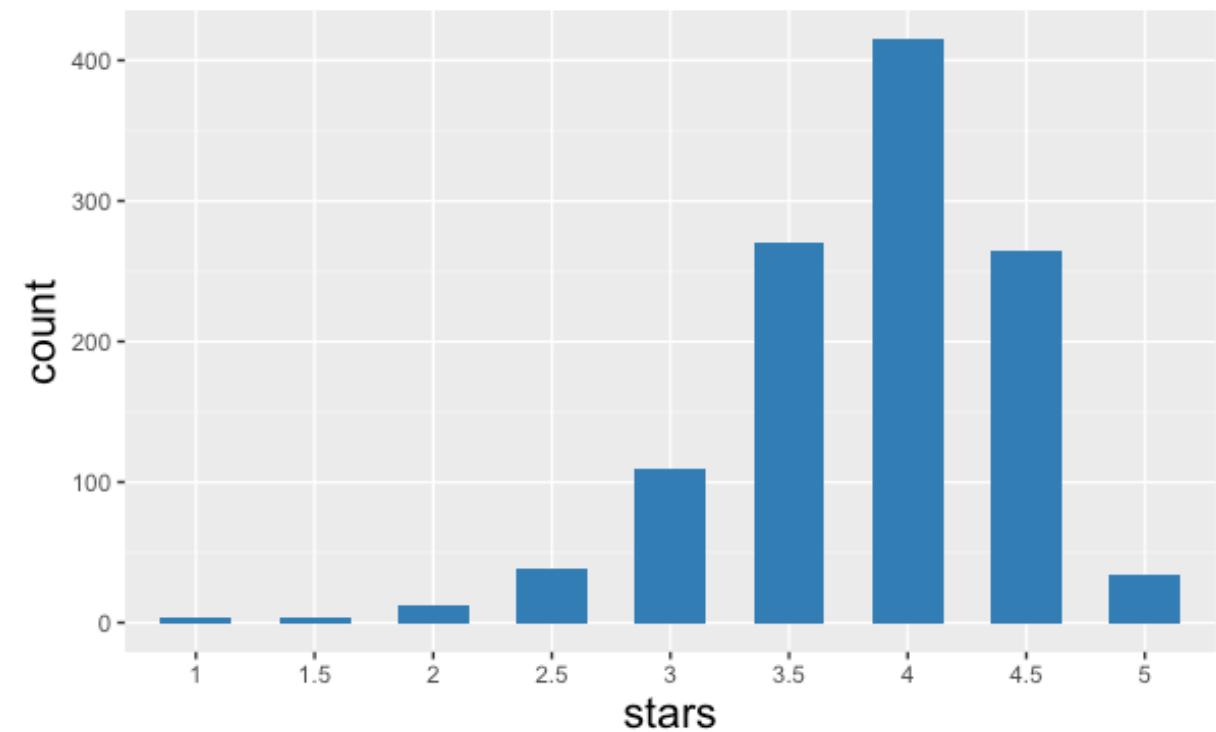


Data Visualization—Attributes (TV)

Has TV



Not has TV



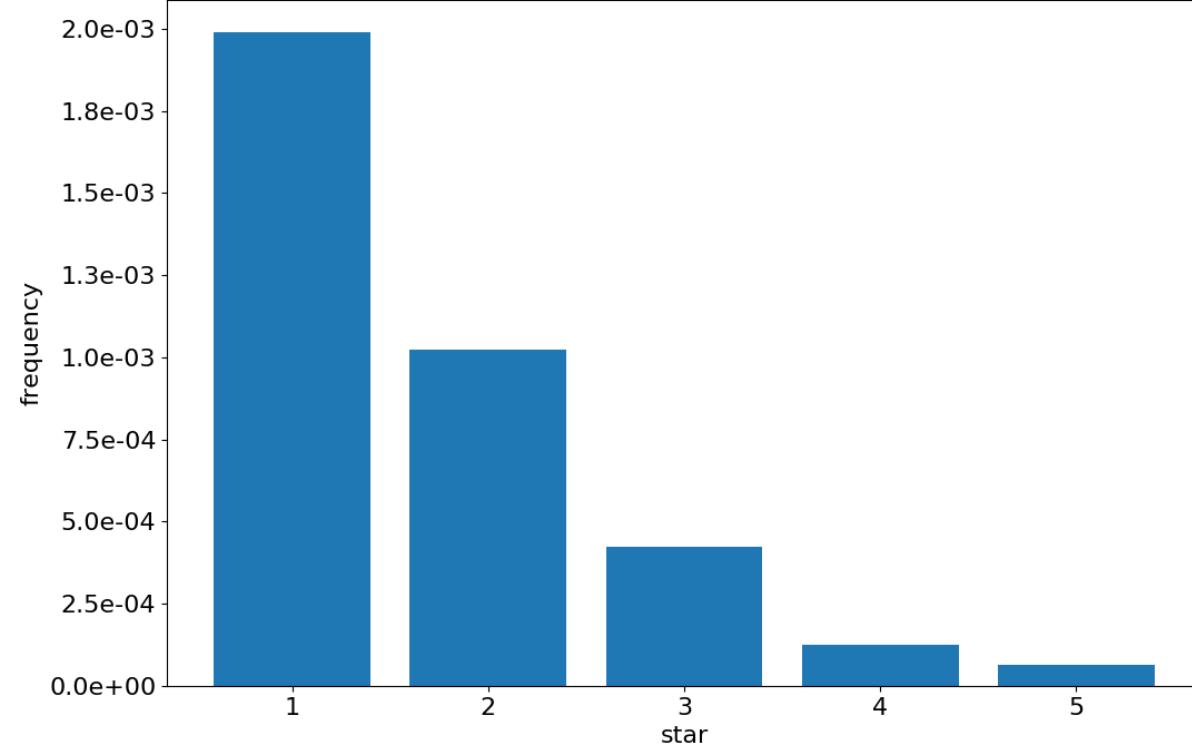
Feature Extraction—TF-IDF

Service		Place		Food			Feeling	
Time	Staff	Atmosphere	Spot	Menu	Price	Drinks	Positive	Negative
Wait	Waitress	Music	Outside	Chicken	Cheap	Beer	Great	Terrible
Table	Friendly	Quiet	Inside	Burger	Not Cheap	Craft Beer	Delicious	Bad
Hour	Attentive	Live	Town	Cheese	Priced	Blood Mary	Nice	Awful
Slow	Rude	Ambience	Street	Salad	Expensive	Old Fashion	Good	Disappointed
Busy	Helpful	Not Clean	Parking	Fries	Affordable	Cocktails	Love	Crowded

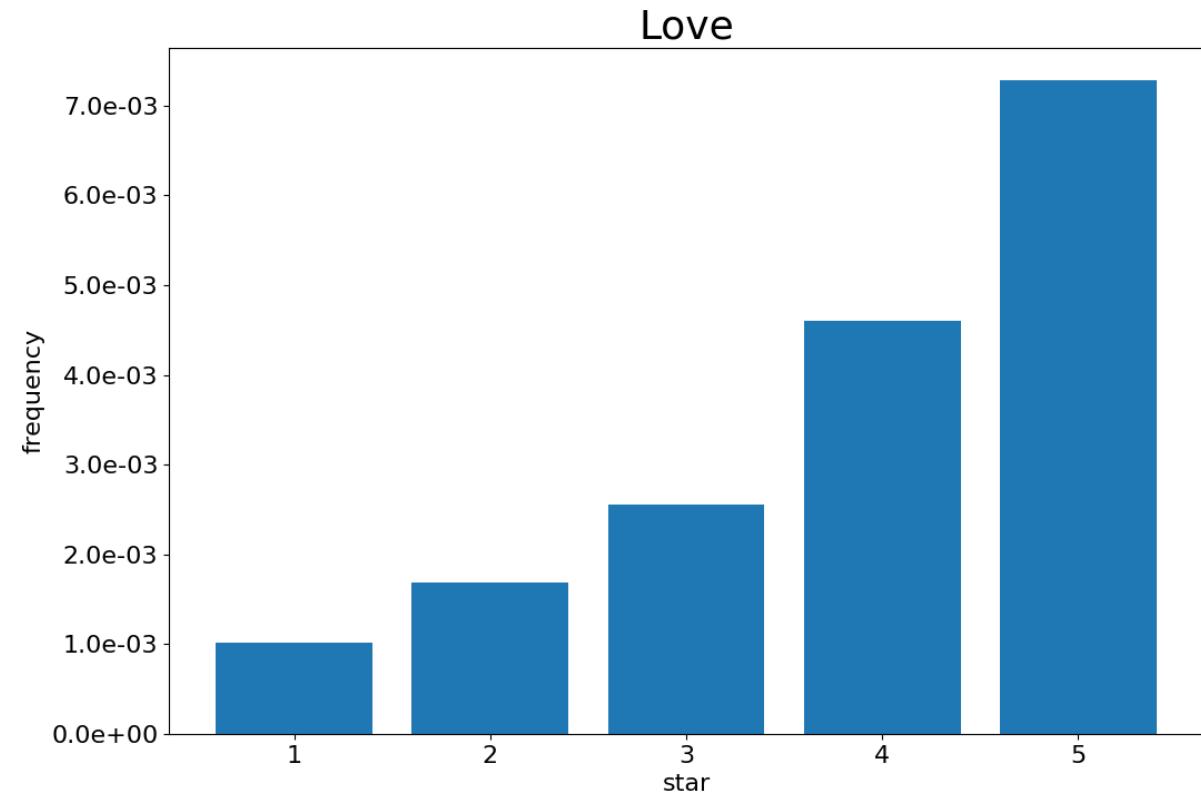


Feature Extraction—TF-IDF

Terrible

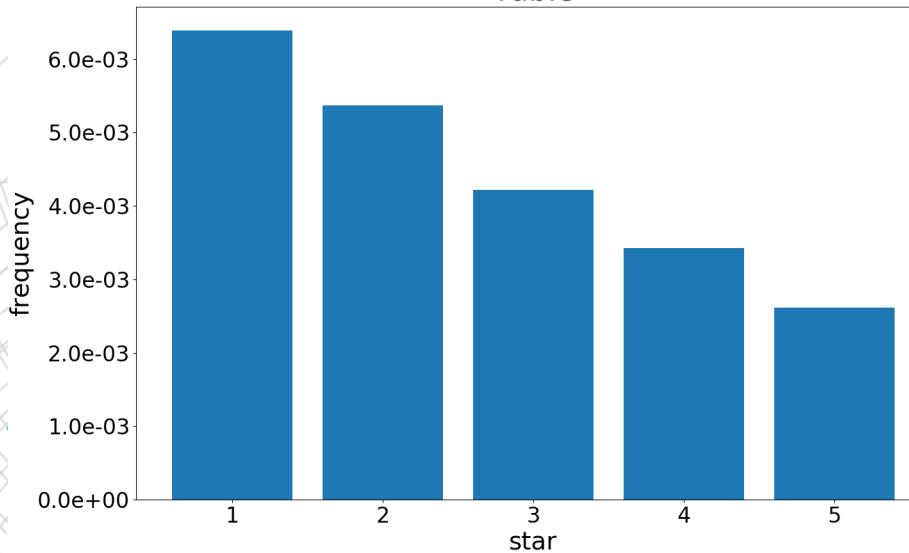


Love

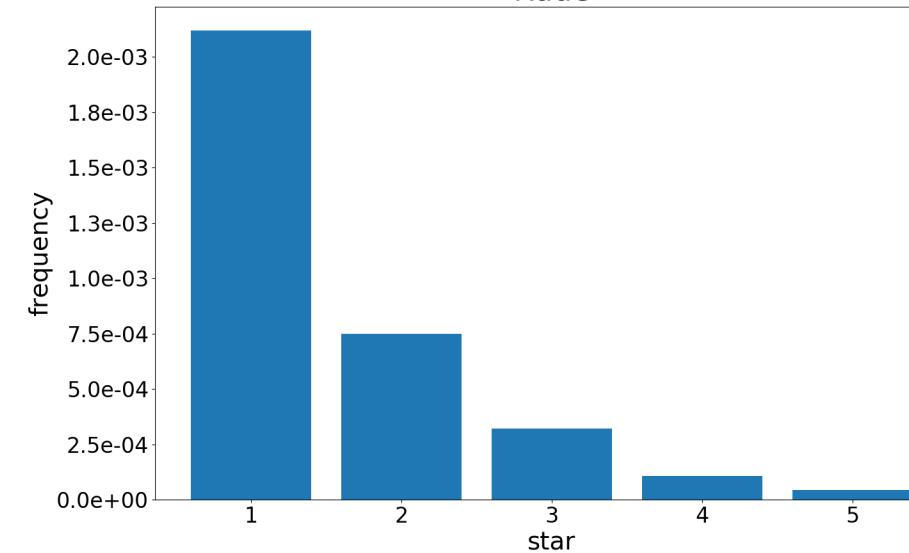


Feature Extraction—TF-IDF

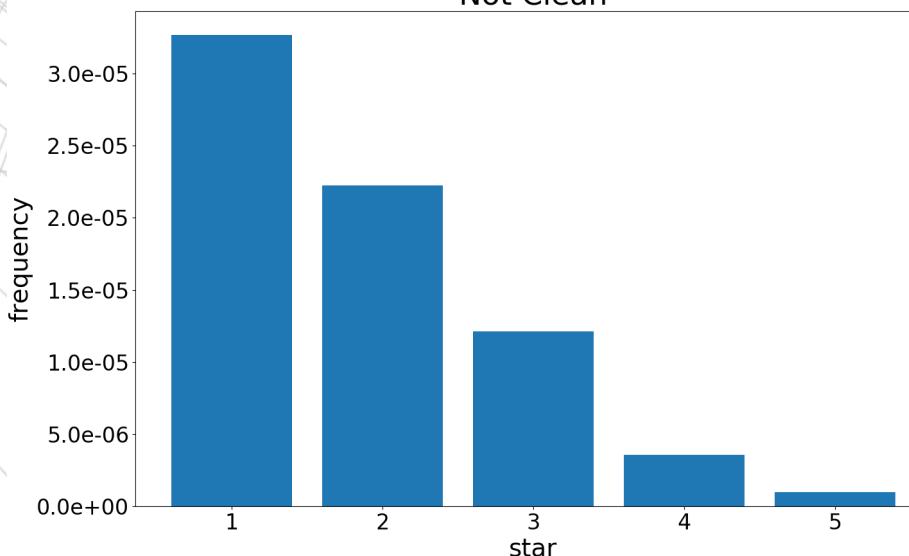
Table



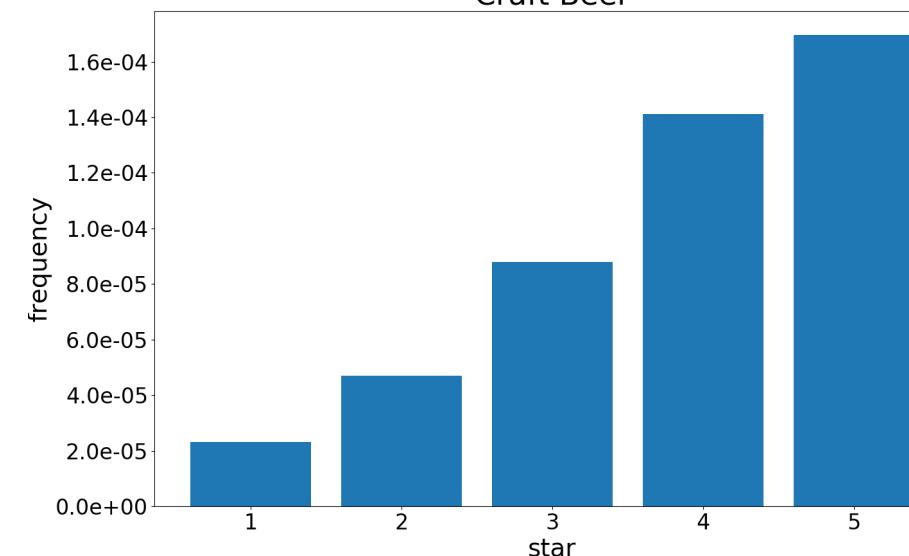
Rude



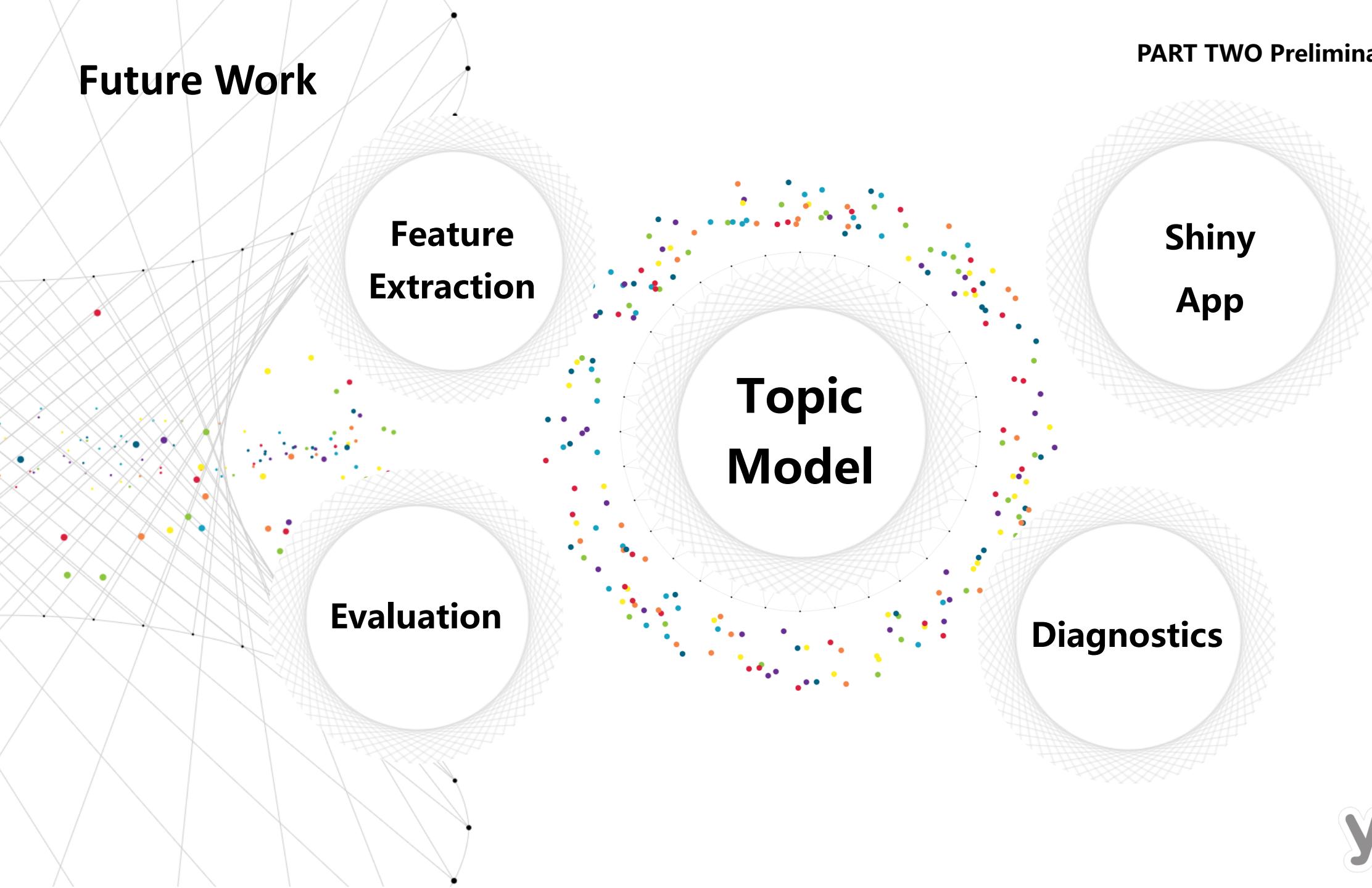
Not Clean

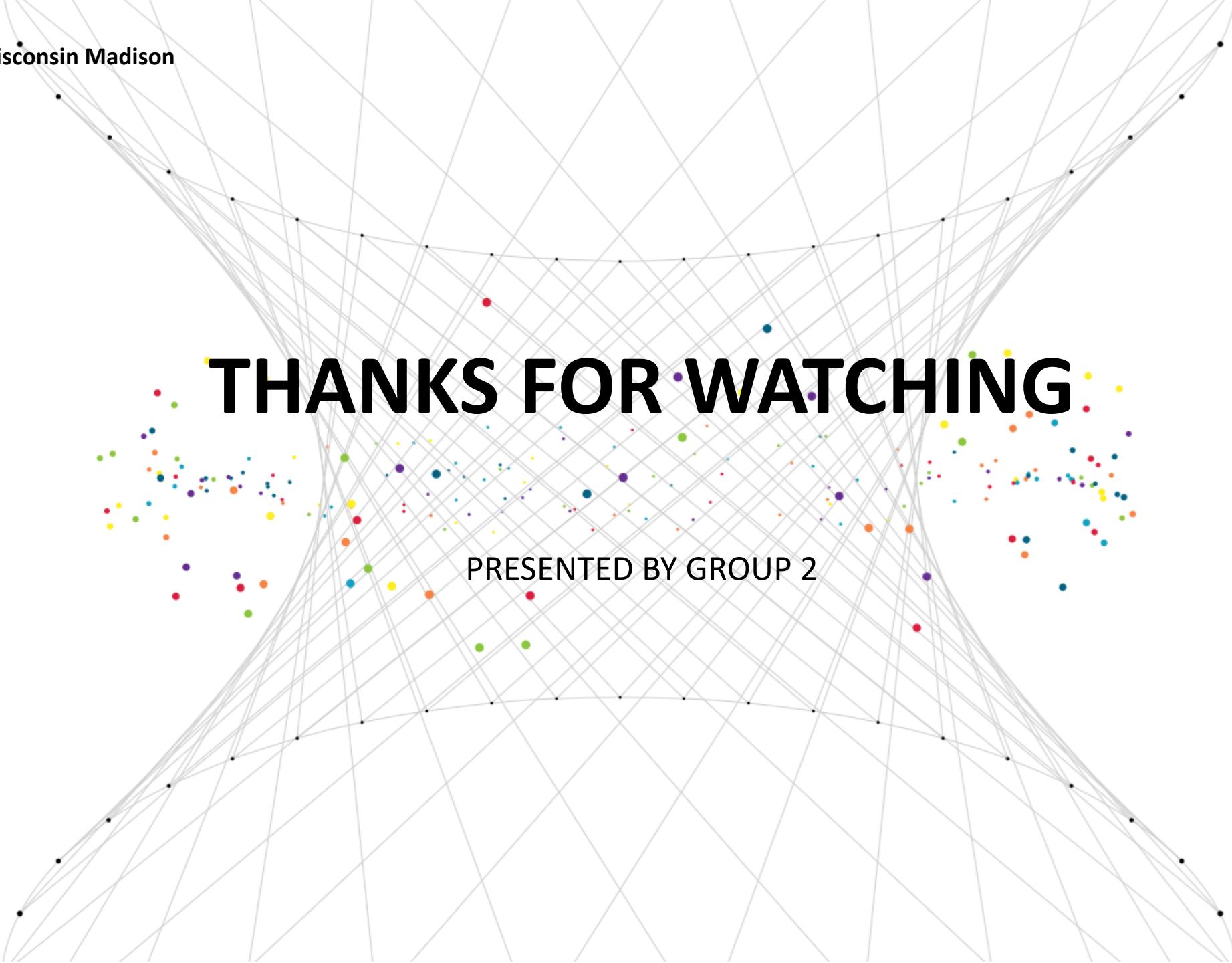


Craft Beer



Future Work





THANKS FOR WATCHING

PRESENTED BY GROUP 2

