

Restaurant Data Analysis

Data 230 - Data Visualization
Data Source: Yelp
Student: Aneshaa Kasula (014558427)

Introduction

Objective

- Build a web service to host an interactive dashboard of any given service and city of the end user's choice
- Leverage publicly available dataset (Yelp) with following details:
 - Businesses
 - Customer Behaviour - Reviews & Ratings or Hourly visits

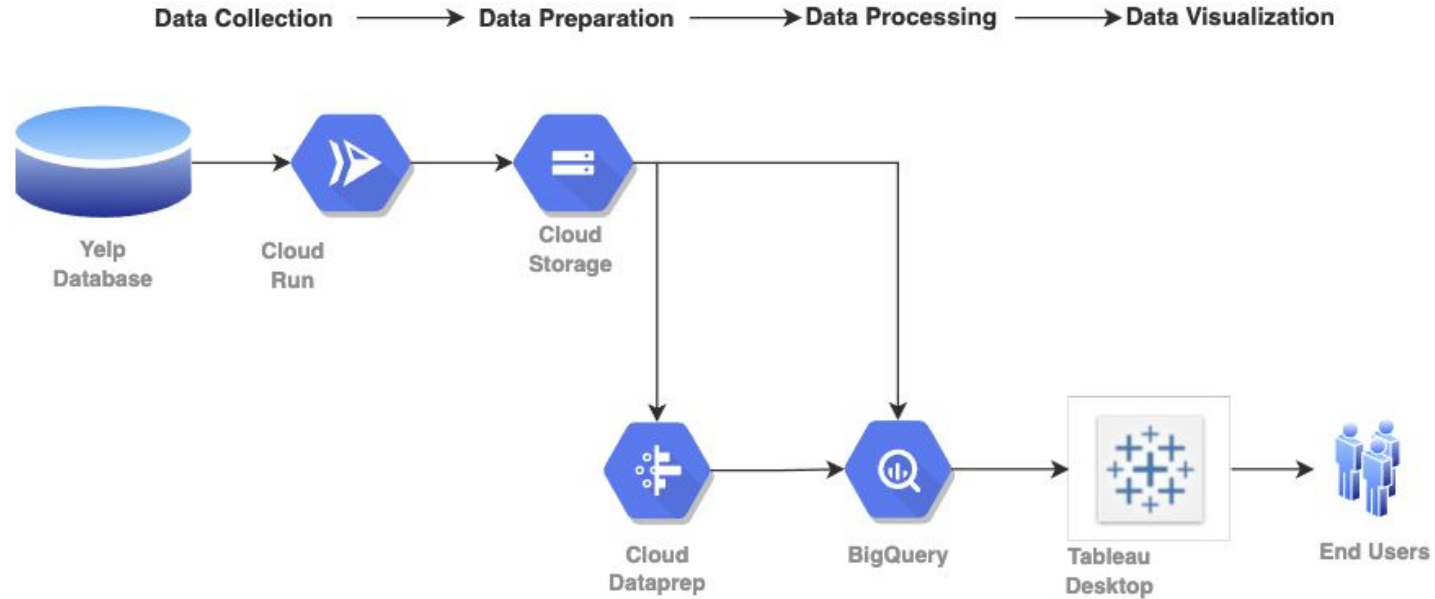
Scope

- Build dashboards with above objectives for end users using Tableau Desktop

Intended Audience

- This project scope caters to 3 types of audiences namely
 - New Investor planning to enter into restaurant business
 - Existing Restaurant Owner and it's chef
 - Restaurant Visitors

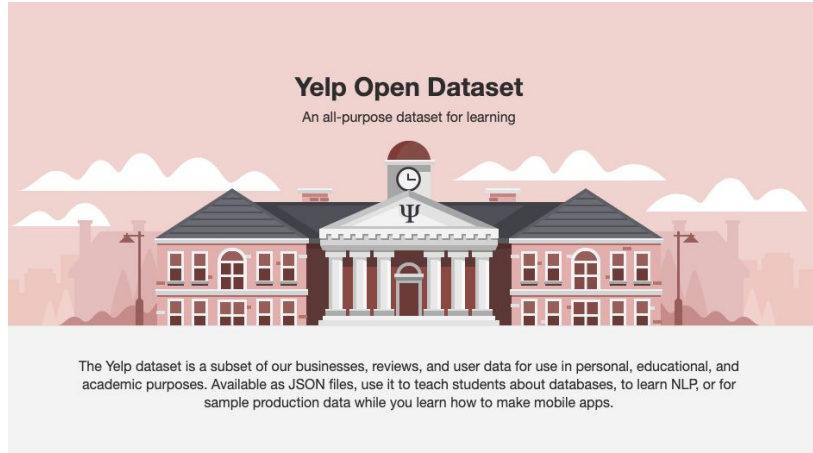
Project Architecture



Project Timeline

Milestone	Expected Date of Completion	Actual Date of Completion
Data Collection	20th September 2021	20th September 2021
Data Preparation	11th October 2021	18th October 2021
Data Processing	1st November 2021	8th November 2021
Data Visualization	15th November 2021	19th November 2021

Data Collection



The Dataset



8,635,403 reviews



160,585 businesses



200,000 pictures



8 metropolitan areas

1,162,119 tips by 2,189,457 users
Over 1.2 million business attributes like hours, parking, availability, and ambience
Aggregated check-ins over time for each of the 138,876 businesses

- Data Source - Yelp
- As it is open dataset with limited data; building dashboards with same
- Ratings and reviews are up to date (2010-2021)
- More than 20 categories of business but restricted to Restaurants as per the scope
 - 115 sub categories
 - Statistics of restaurants data in USA
 - 43283 restaurants registered to Yelp
 - 14 states
 - 433 cities
 - 604 Postal codes
- Leveraged two subsets of dataset:
 - Businesses
 - File : business.json
 - Check-ins
 - File : checkin.json

Dataset Link: <https://www.yelp.com/dataset>

Data Preparation & Processing

```
{
  "business_id": "61Yb2HFDym3zjuRg0shjw",
  "name": "Oskar Blues Taproom",
  "address": "921 Pearl St",
  "city": "Boulder",
  "state": "CO",
  "postal_code": "80302",
  "latitude": 40.0175444,
  "longitude": -105.2833481,
  "stars": 4.0,
  "review_count": 86,
  "is_open": 1,
  "attributes": {
    "RestaurantsTableService": "True",
    "WiFi": "u'free'",
    "BikeParking": "True",
    "BusinessParking": "{ 'garage': False, 'street': True",
    "BusinessAcceptsCreditCards": "True",
    "RestaurantsReservations": "False",
    "WheelchairAccessible": "True",
    "Caters": "True",
    "OutdoorSeating": "True",
    "RestaurantsGoodForGroups": "True",
    "HappyHour": "True",
    "BusinessAcceptsBitcoin": "False",
    "RestaurantsPriceRange2": "2",
    "Ambience": "{ 'touristy': False, 'hipster': False",
    "HasTV": "True",
    "Alcohol": "beer_and_wine",
    "GoodForMeal": "{ 'dessert': False, 'latenight': F",
    "DogsAllowed": "False",
    "RestaurantsTakeOut": "True",
    "NoiseLevel": "u'average'",
    "RestaurantsAttire": "casual",
    "RestaurantsDelivery": "None"
  },
  "categories": "Gastropubs, Food, Beer Gardens, Resta
```

```
{
  "business_id": "—0zrn43LEaB4jUWTQH_Bg",
  "date": "2010-10-08 22:21:20, 2010-11-01 21:29:14, 2010-12-23 22:55:45, 2011-04-08 17:14:59, 2011-04-11 21:28:45,
  "business_id": "—164t1nclzmc7eD1JMW",
  "date": "2010-02-26 02:06:53, 2010-02-27 08:00:09, 2010-03-04 02:00:59, 2010-03-11 01:24:46, 2010-03-17 02:29:17,
```

aneshaak_yelp			
Location	Storage class	Public access	Protection
us-west1 (Oregon)	Standard	Not public	None
OBJECTS	CONFIGURATION	PERMISSIONS	PROTECTION
Buckets > aneshaak_yelp			
UPLOAD FILES UPLOAD FOLDER CREATE FOLDER MANAGE HOLDS DOWN			
Filter by name prefix only Filter Filter objects and folders			
<input type="checkbox"/>	Name	Size	
<input type="checkbox"/>	yelp_academic_dataset_business.json	116.1 MB	
<input type="checkbox"/>	yelp_academic_dataset_business_restaurant.json	89 MB	
<input type="checkbox"/>	yelp_academic_dataset_checkin.json	379.8 MB	

restaurant_table

SCHEMA

DETAILS

PREVIEW

Row

business_id

business_type

name

address

city

state

postal_code

latitude

longitude

rating

review_count

is_open

alias_categories

Category

1

v-fP4Cg7DQ9oaDx7xJzQ

Restaurants

McDonald's

15359 I 35

Buda

TX

78610

30.0890987283

-97.8173089027

1.0

34

1

Restaurants, Food, Coffee & Tea, Burgers, Fast Food

Burgers

2

zLwJZM2V1hecrWoOykhA

Restaurants

Subway

15300 S I h 35, Ste 380

Buda

TX

78610

30.0869967713

-97.8214660086

1.0

20

1

Restaurants, Sandwiches, Fast Food

Fast Food

3

CIKJLH4WuXgB-mnf-gbUw

Restaurants

Long John Silvers

720 Old San Antonio Rd

Buda

TX

78610

30.0875883

-97.8224388

1.0

7

1

Fast Food, Chicken Wings, Restaurants, Seafood

Chicken Wings

4

3GBd_LC2gYDgS9JMqfQ

Restaurants

Taco Bell

265 Washington Street

Stoughton

MA

2072

42.13886

-71.10265

1.0

15

1

Tex-Mex, Tacos, Mexican, Fast Food, Restaurants

5

bA8CokKVRJIEIC_x8MqYA

Restaurants

Papa John's Pizza

1740 Semoran Boulevard

Winter Park

FL

32792

28.619656

-81.312359

1.0

7

1

Pizza, Restaurants

6

9_HCOc_LJX86yzeE2SLjg

Restaurants

Sonic Drive-In

5110 Ross Rd

Del Valle

TX

78617

30.1800268

-97.6145246

1.0

19

1

Burgers, Food, Ice Cream & Frozen Yogurt

res_checkin_byid_norm

SCHEMA

DETAILS

PREVIEW

res_checkin_byid_norm

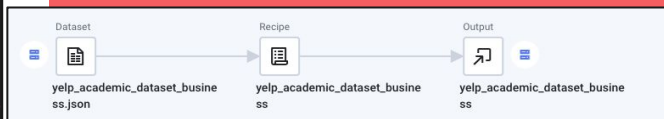
SCHEMA

DETAILS

PREVIEW

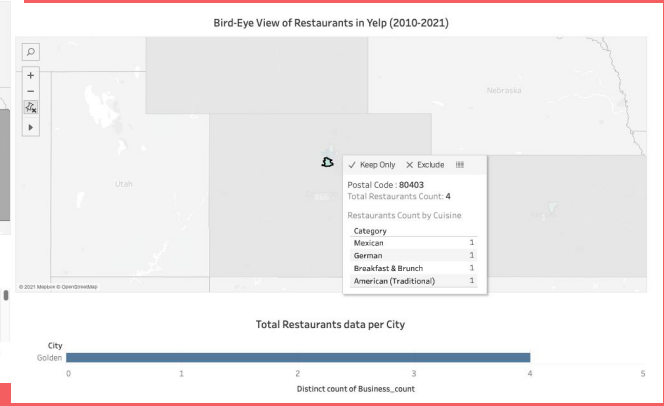
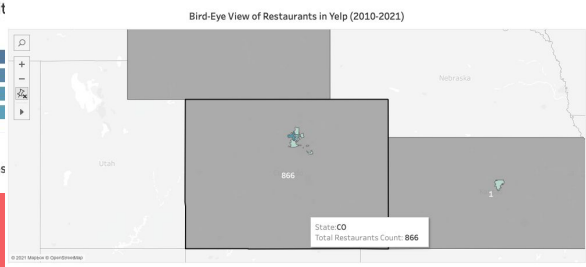
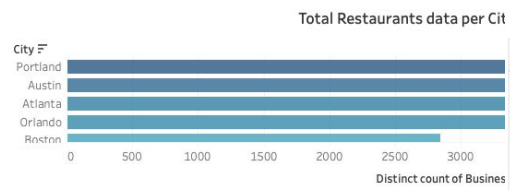
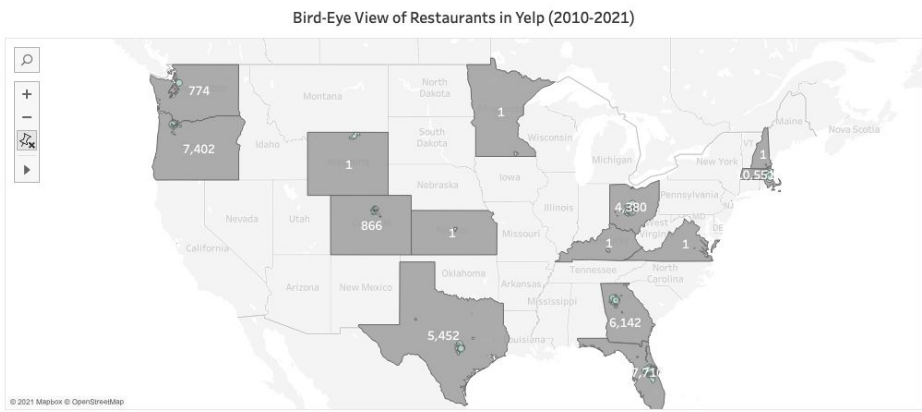
Row	date	business_id
1	2016-07-25 00:19:58	YpYnDN1U6vz8_3XhLcBgWQ
2	2020-08-26 12:48:40	YpYnDN1U6vz8_3XhLcBgWQ
3	2016-06-29 02:05:42	aq7aekhLkhy-ploNm1IDSg
4	2017-02-10 02:53:33	aq7aekhLkhy-ploNm1IDSg
5	2018-05-02 04:11:00	aq7aekhLkhy-ploNm1IDSg

- Business Data:
 - In GCP DataPrep,
 - Removed unnecessary columns such as hours, additional attributes
 - Handled Mismatched values
 - ~1% Missing values are removed
 - In Bigquery, created table with updated data and it is used for building dashboards.
- Checkin's Data:
 - From cloud storage, created table in Bigquery with date column normalized from multiple values to single values.



Dashboard 1 - GeoSpatial (2010-2021)

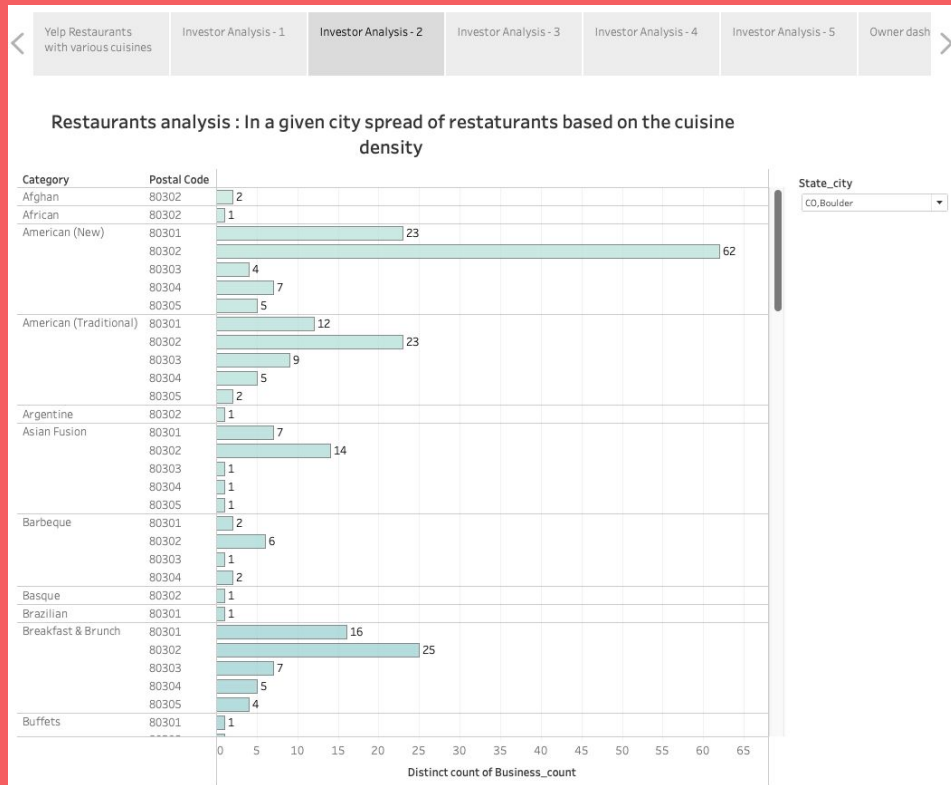
- First part of dashboard shows restaurants with different types of cuisines
 - Count of restaurants per state
 - Drill down to Postal code
 - Count restaurants
 - Top Ratings with restaurants name
- Second part shows number of business per city





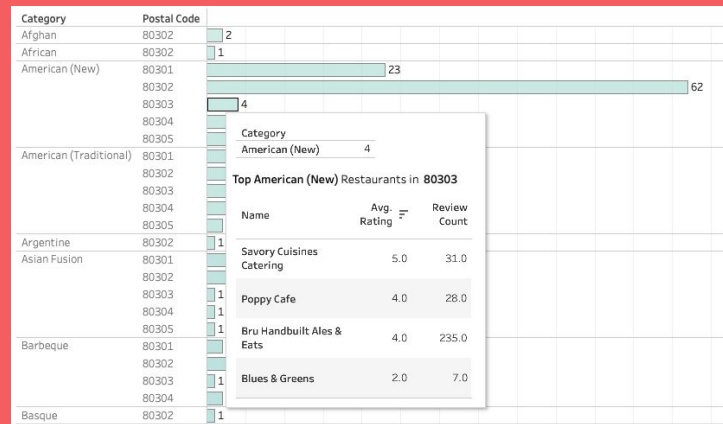
Dashboard 2 - Heatmap

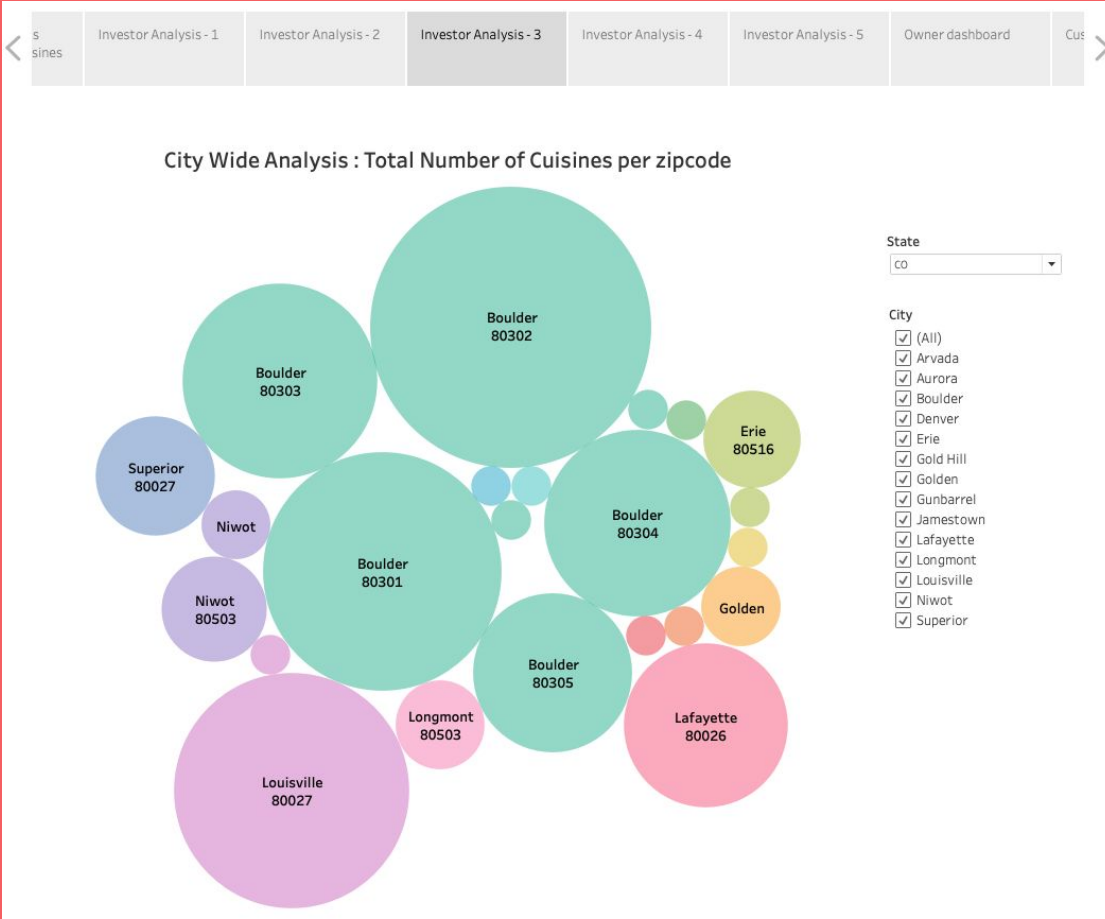
- Investor Analysis
 - Uses filter state and city
 - Get info on demand of restaurants per postal code across different cuisines
 - More demand then Open new business neighborhood postal areas
 - No businesses but people are migrating then open businesses of particular cuisine



Dashboard 3 – Bar chart

- Investor gets a glimpse on number of existing restaurants with names of cuisines for a given city





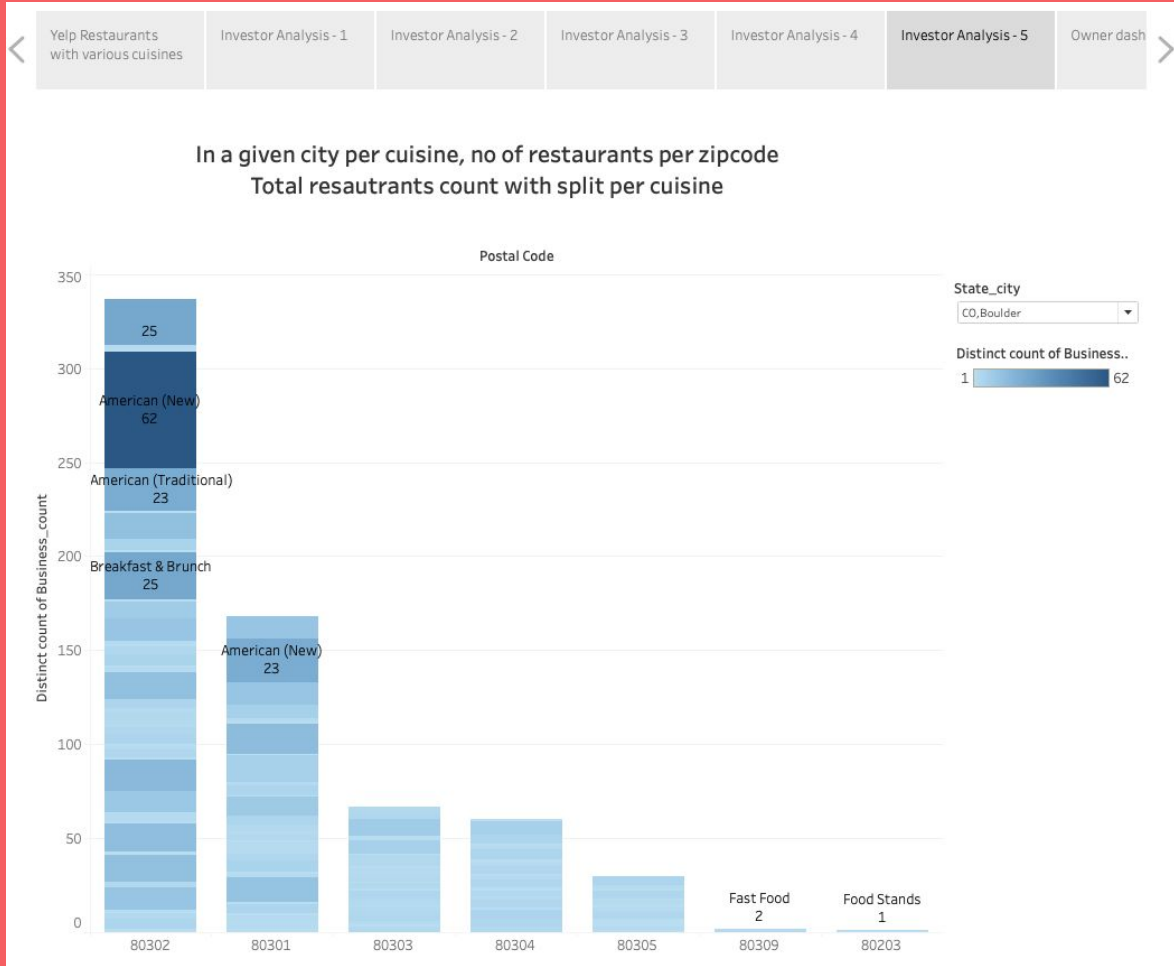
Dashboard 4 - Bubbles chart

- High level view for an Investor
 - Demand for restaurants in city per postal area



Dashboard 5 – Line chart

- High level view for an Investor
 - People visits from 2010 to 2021
 - Drastic decline in 2020 due to Covid-19 pandemic



Dashboard 6 – Bar Chart

High level view for an Investor

- Filter city
- Restaurants per postal area



Dashboard 7 - Heat Map



Restaurant Owner

Filter

- Year
- Month
- City
- Restaurant Name
- Darker color indicate busy hours
- Arrange raw materials, staff and cutlery
- For instance,
 - Morning hours
 - Plan ahead before day
 - Weekends
 - Plan ahead during early hours

In a Given City, Customers could analyze through all the restaurants in that city across different cuisines and decide on a restaurant based on it's Review Count and Average Rating

State: CO

City: Boulder

- [illegible]



Dashboard 9 – Heat Map

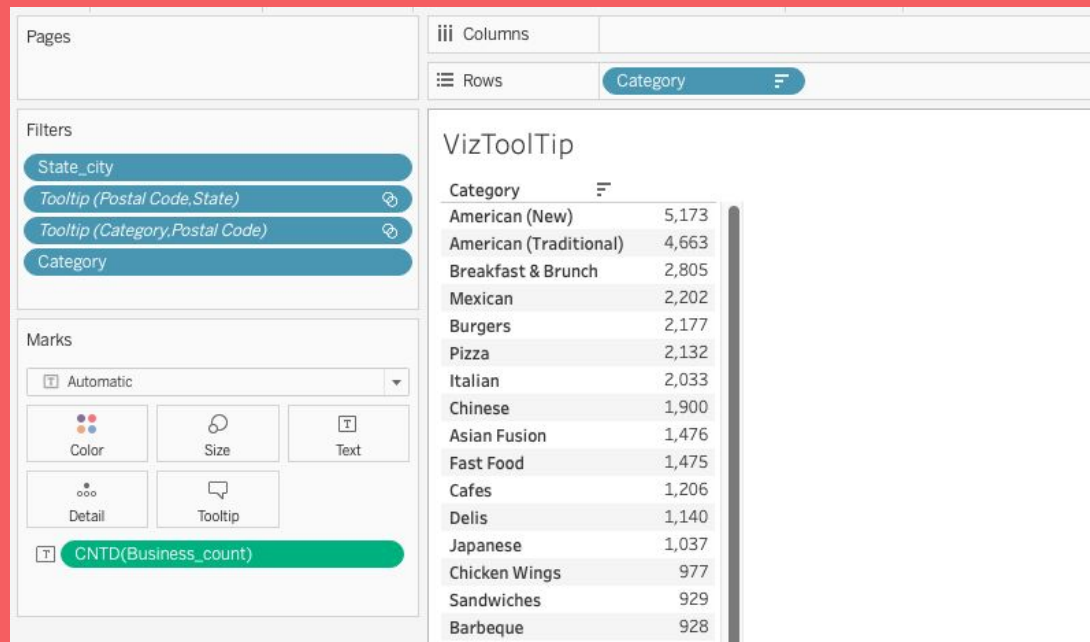
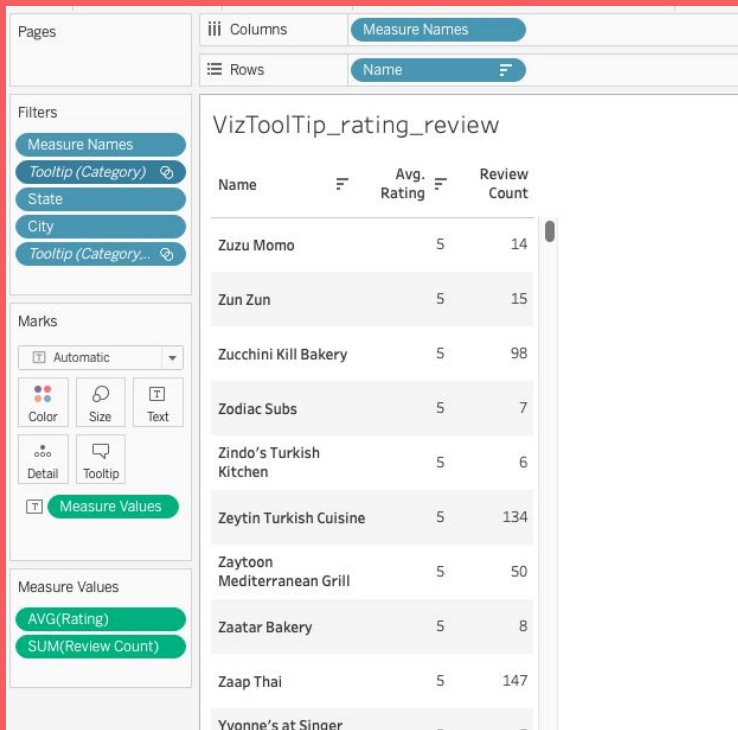
High level view for Customer

Filter

- Year (Recent)
- Month (Seasonal impact)
- State and City
- Choice of Restaurant
- Hourly demand per day

Tabular Viz in Tooltips

- Restaurants with Ratings



Future Work

- Imbibe Google Data also for details
- Build comparison dashboards of various businesses from the data sourced from Yelp and Google.
- The current scope is limited to restaurants but the objective is to spawn out across different categories in the market such as grocery stores, arts and crafts, salons etc.
- Dynamic web page, with active data collection from all the data sources using API calls.

Thank You!