

Project 1: Data Analysis of Restaurants in Washington DC

...

Amy Sylla & Oluwatobi Akinsanya
September 8, 2020

Outline:

A client wants to open a restaurant in Washington DC. What type of restaurant should they open and where in DC ?

Our Analysis will answer **4 questions**:

- What is the most competitive restaurant type in each DC Quadrant?
- What is the average restaurant quality for restaurant types in each DC Quadrant?
- What are the best graded restaurants in DC ? Do they reflect their demographics ?
- Food trucks vs Restaurants, what to pick ?

What is the most competitive restaurant type in each DC Quadrant?

This data will help us find out which restaurant type may have been exhausted in DC and which restaurant types have opportunity to grow



What is the average/best restaurant quality for restaurant types in each DC Quadrant?

This data will help identify if there is a particular restaurant type that may be associated with lower food quality. If so, this may present an opportunity to stand out.



Do ethnic groups in DC impact restaurant types in each DC Quadrant?

With this question, we can identify any trends that may be related to DC's ethnicity. This will help us understand if a particular group of people impact the amount of a restaurant type in DC. If there is a trend, the new restaurant may have to serve those groups of people in some way with the new restaurant.



Understanding the problem: Methods used

What is needed?

- Total number of restaurants in DC
- Ratings of Restaurants type ratings
- Description of ethnic groups in DC

How do we find that?

- Find a good data source
- Determine quadrants
- Zip codes/Coordinates

Treatment + Analysis

- Loading data
- Combining data
- Plot data
- Analyse

Sources of Informations

Retrieving data: Food

Yelp



Pros:

- Detailed
- Developer friendly
- Includes both Api and CSV files

Cons:

- No format
- Only allows 1000 data maximum per entry

Retrieving data: Population

United States Census



Pros:

- Very detailed
- Trusted

Cons:

- Cryptic Api calls
- Time consuming
- Does not include good coordinates for lower state level

Retrieving data: Population

DC Health

DC | **HEALTH**
GOVERNMENT OF THE DISTRICT OF COLUMBIA

- Very detailed
- Trusted
- CSV files included

- Very detailed
- Trusted
- CSV files included

Processing data

Loading data: Json and CSV

Json

```
#Call
data = json.loads(req.text)
print(json.dumps(data, indent=4))
```

```
{
  "businesses": [
    {
      "id": "6lBhCjv-imBpa3IYKVSCCA",
      "alias": "maxs-kosher-grill-washington",
      "name": "Max's Kosher Grill",
      "image_url": "https://s3-media4.fl.yelpcdn.com/bphoto/YF130-FV2LuRLy3vueGaCA/o.jpg",
      "is_closed": false,
      "url": "https://www.yelp.com/biz/maxs-kosher-grill-washington?adjust_creative=1m4MG5a37UVgnQoLAVWITw&utm_campaign=yelp_api_v3&utm_medium=api_v3_business_search&utm_source=1m4MG5a37UVgnQoLAVWITw",
      "review_count": 5,
      "categories": [
        {
          "alias": "kosher",
          "title": "Kosher"
        },
        {
          "alias": "falafel",
          "title": "Falafel"
        }
      ]
    }
  ]
}
```

Loading data: Json and CSV

CSV

#Southwest Ethnicity Data

```
zip_20024_Hisp = pd.read_csv('DC_Health_Data/Hispanic_Latino_Population_by_Race_Zip_Code_20024.csv')  
zip_20024_Non = pd.read_csv('DC_Health_Data/Non_Hispanic_Latino_by_Race_Zip_Code_20024.csv')
```

```
zip_20032_Hisp = pd.read_csv('DC_Health_Data/Hispanic_Latino_Population_by_Race_Zip_Code_20032.csv')  
zip_20032_Non = pd.read_csv('DC_Health_Data/Non_Hispanic_Latino_by_Race_Zip_Code_20032.csv')
```

#Southeast Data

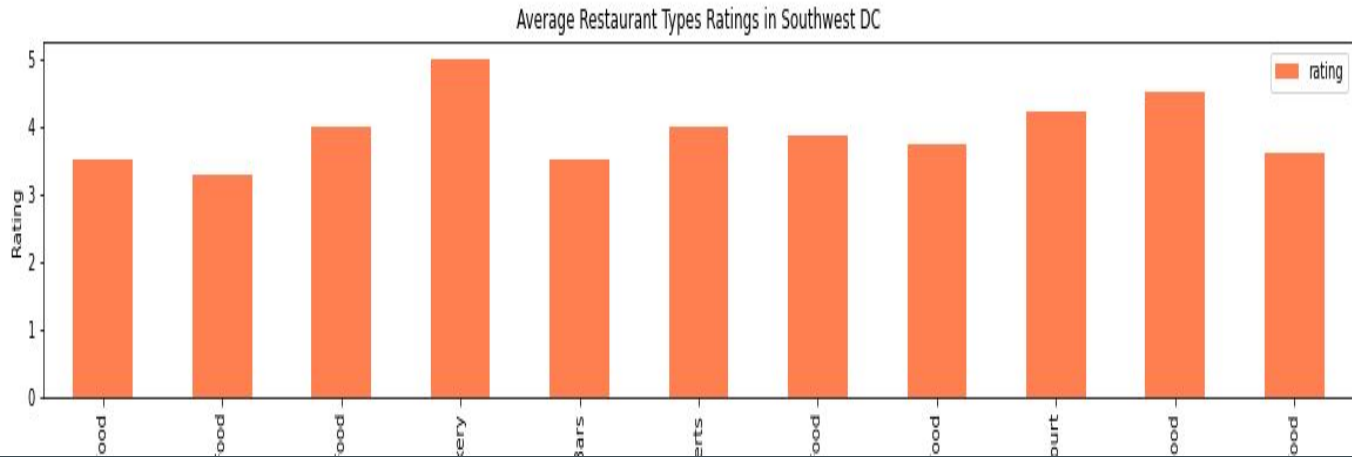
```
zip_20003_Hisp = pd.read_csv('DC_Health_Data/Hispanic_Latino_Population_by_Race_Zip_Code_20003.csv')  
zip_20003_Non = pd.read_csv('DC_Health_Data/Non_Hispanic_Latino_by_Race_Zip_Code_20003.csv')
```

```
Ward_8_Hisp = pd.read_csv('DC_Health_Data/Hispanic_Latino_Population_by_Race_Ward_Ward_8.csv')  
Ward_8_Non = pd.read_csv('DC_Health_Data/Non_Hispanic_Latino_by_Race_Ward_Ward_8.csv')
```

Framing the data: Pandas

	name	category	latitude	longitude	rating	address	State	Zipcode	initial
0	Surfside	Tacos	38.943537	-77.077791	3.5	4200 Wisconsin Ave NW	DC	20015	NW
2	Tabla	Georgian	38.931088	-77.023394	5.0	3227 Georgia Ave NW	DC	20010	NW
3	Viet Chopsticks	Vietnamese	38.945732	-77.064860	4.5	4304 Connecticut Ave NW	DC	20008	NW
5	Rosemary Bistro Cafe	French	38.955090	-77.070230	5.0	5010 Connecticut Ave NW	DC	20008	NW
6	Uptown Market	Grocery	38.947138	-77.064944	4.5	4465 Connecticut Ave NW	DC	20008	NW
...
934	Po Boy Jim	Bars	38.916570	-77.024275	3.5	1934 9th St NW	DC	20001	NW
936	Pi Pizzeria	Pizza	38.897154	-77.024709	4.0	910 F St NW	DC	20004	NW
940	LiLLiES Restaurant & Bar	Mediterranean	38.928427	-77.054125	3.5	2915 Connecticut Ave NW	DC	20008	NW
941	Angelico Pizzeria	Pizza	38.948699	-77.079876	3.5	4529 Wisconsin Ave NW	DC	20016	NW
947	Cafe Saint-Ex	American (New)	38.915423	-77.031673	3.5	1847 14th St NW	DC	20009	NW

Plotting Data:



Analysis: Pandas, logic and a lot of courage

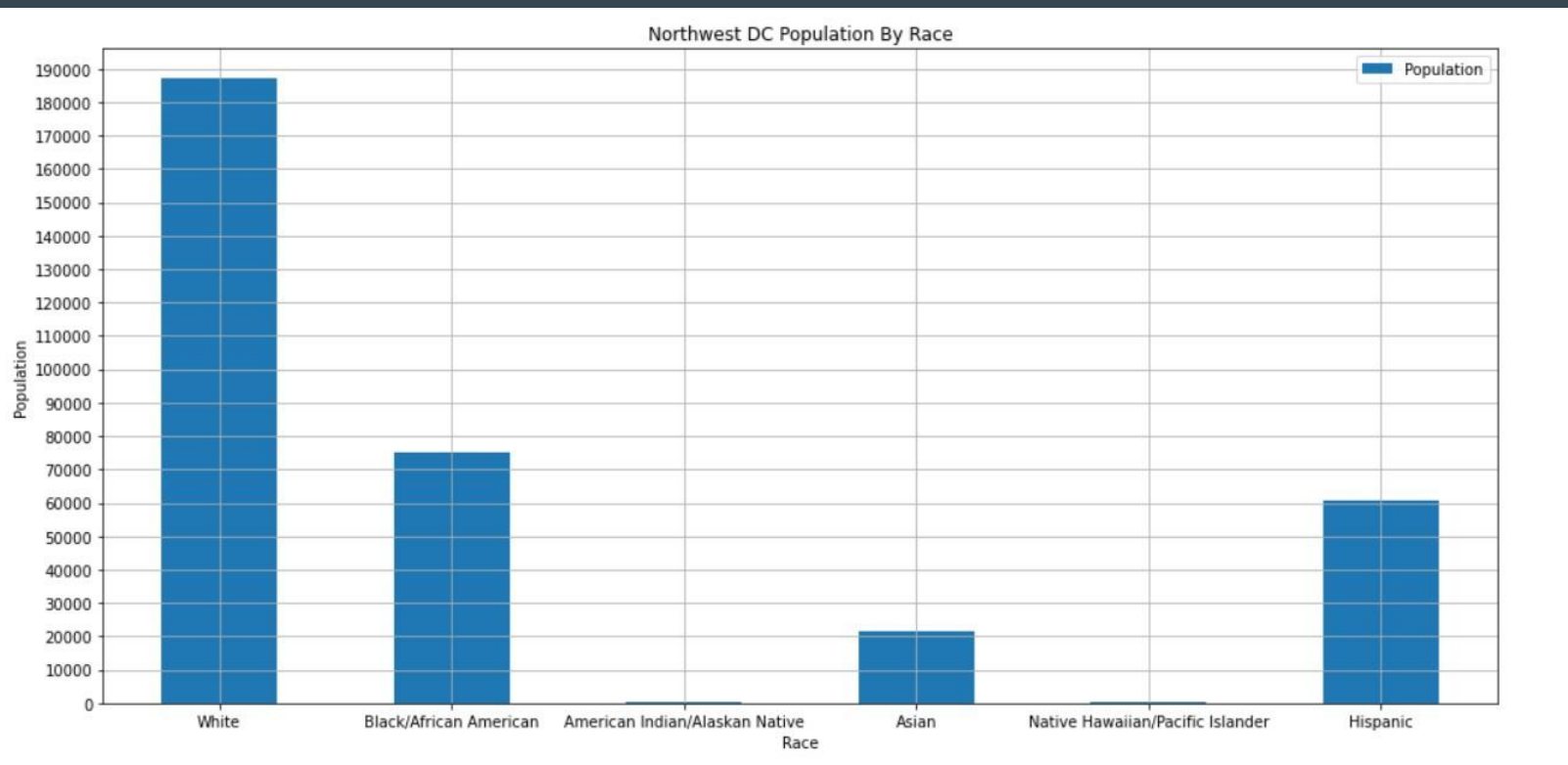
	rating
count	20.000000
mean	4.078740
std	0.275229
min	3.500000
25%	3.948005
50%	4.067254
75%	4.220395
max	4.600000

American Food	169
Asian Food	142
European Food	112
Fast Food	91
Latino Food	77
Bars	61
Coffee & Tea	38
Seafood	33
Food Court	24
African Food	19
Breakfast & Brunch	15
Bakery	13
Desserts	13
Middle Eastern	12
Caribbean	10
Vegetarian	5
Hawaiian Food	5
Vegan	4
Cajun/Creole	3
Trinidadian	1

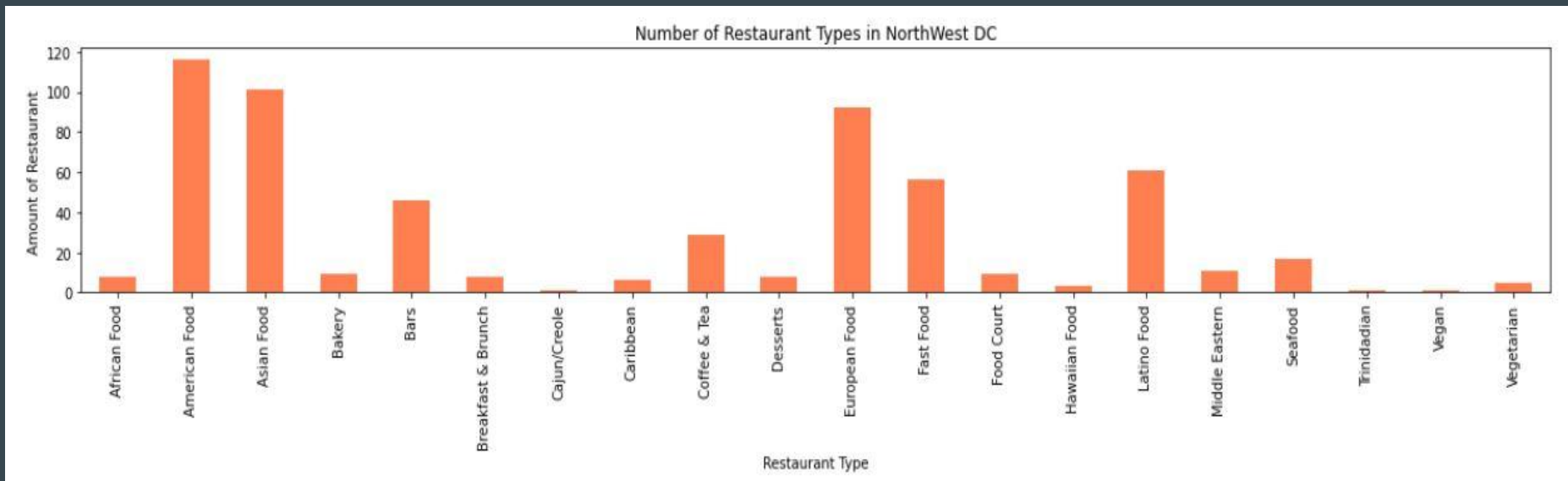
Name: category, dtype: int64

Results

Northwest

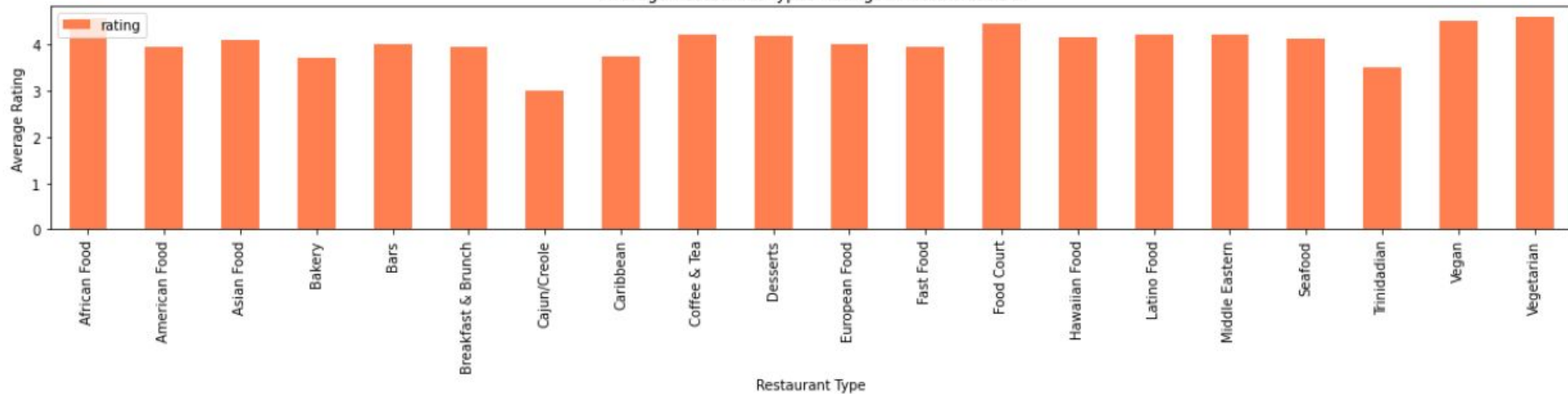


Northwest

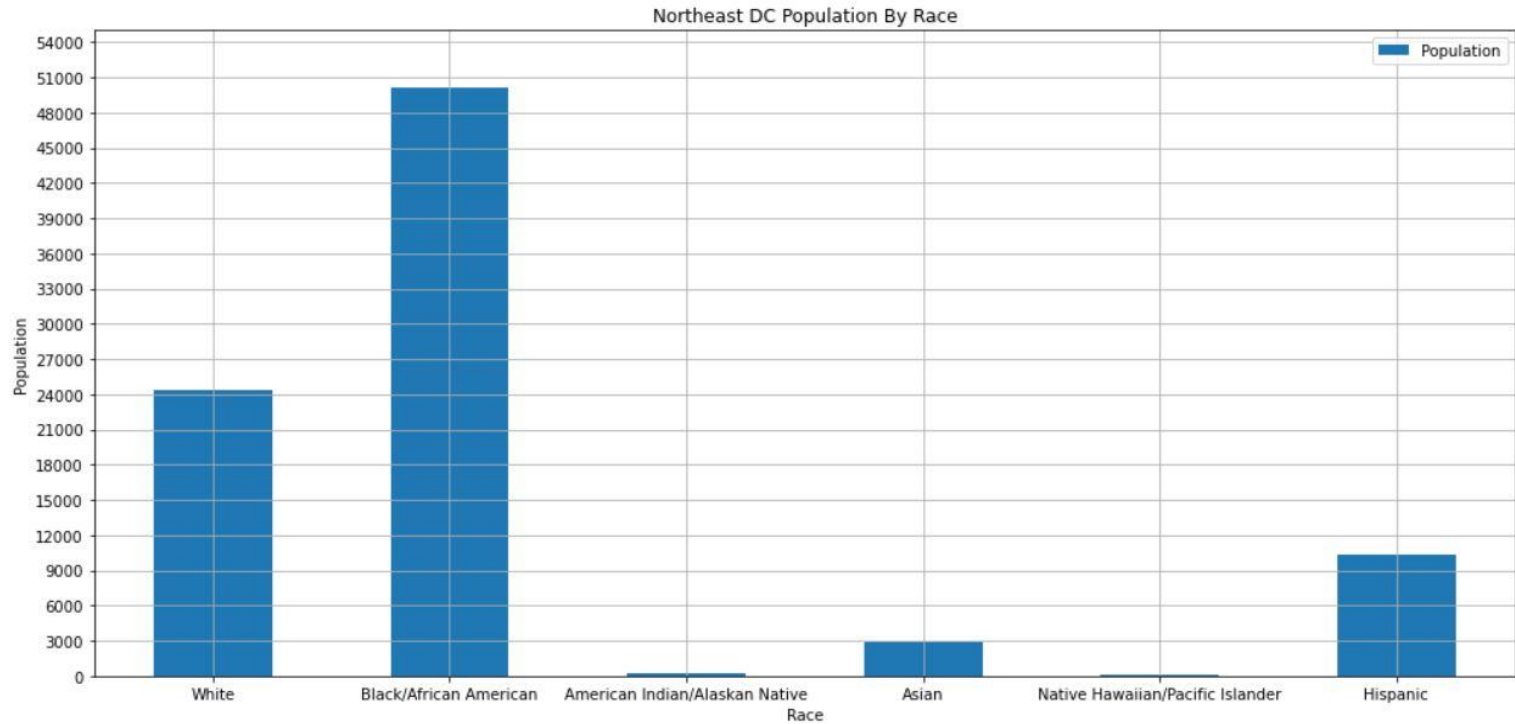


Northwest

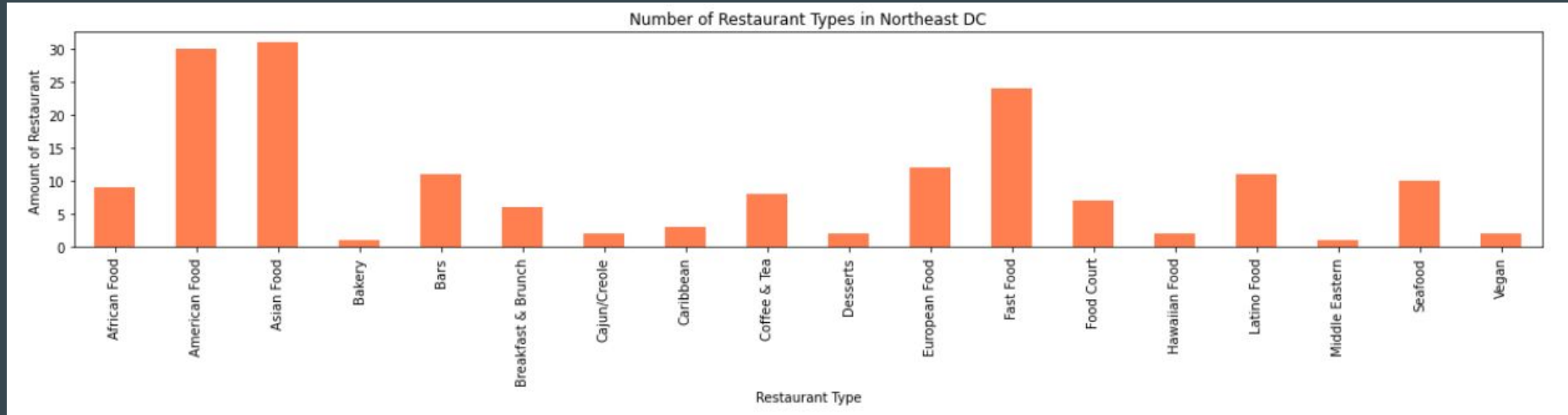
Average Restaurant Types Ratings in Northwest DC



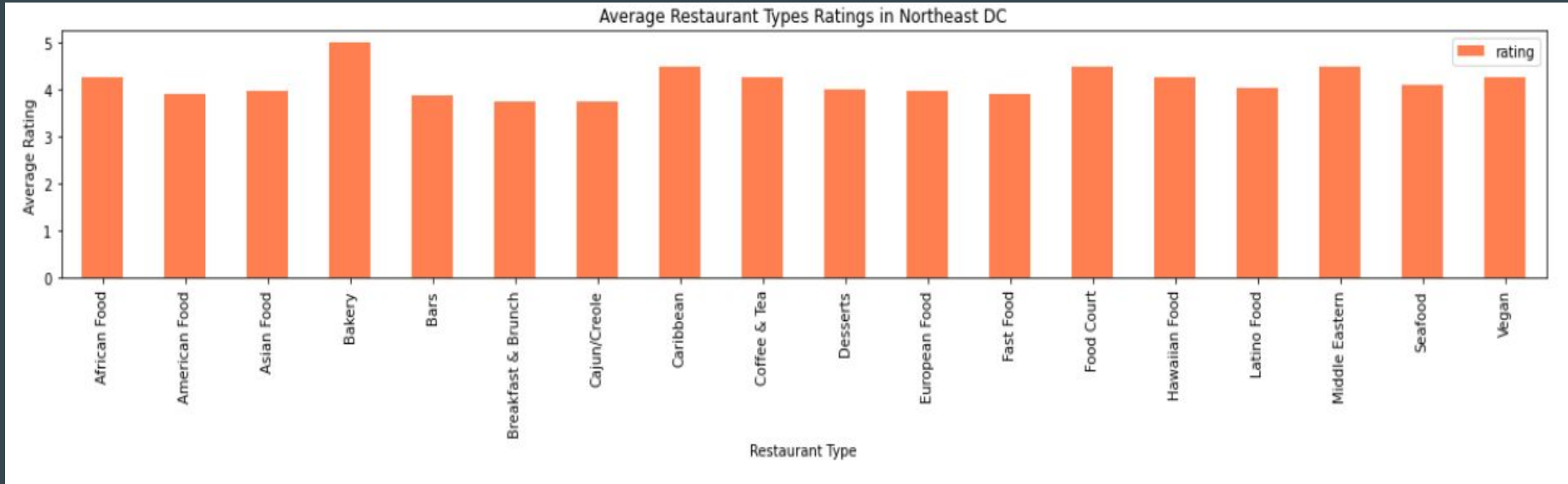
Northeast



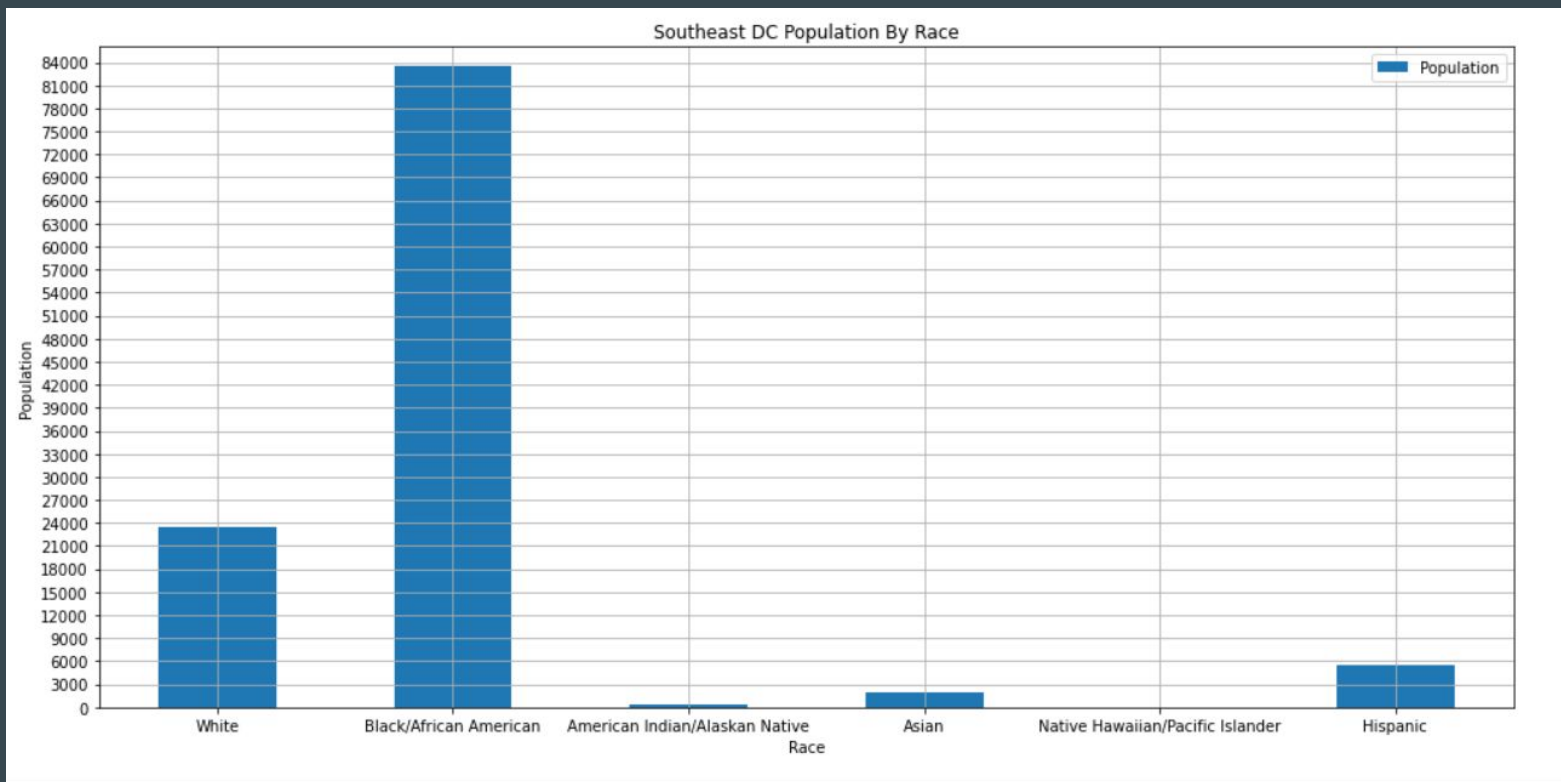
Northeast



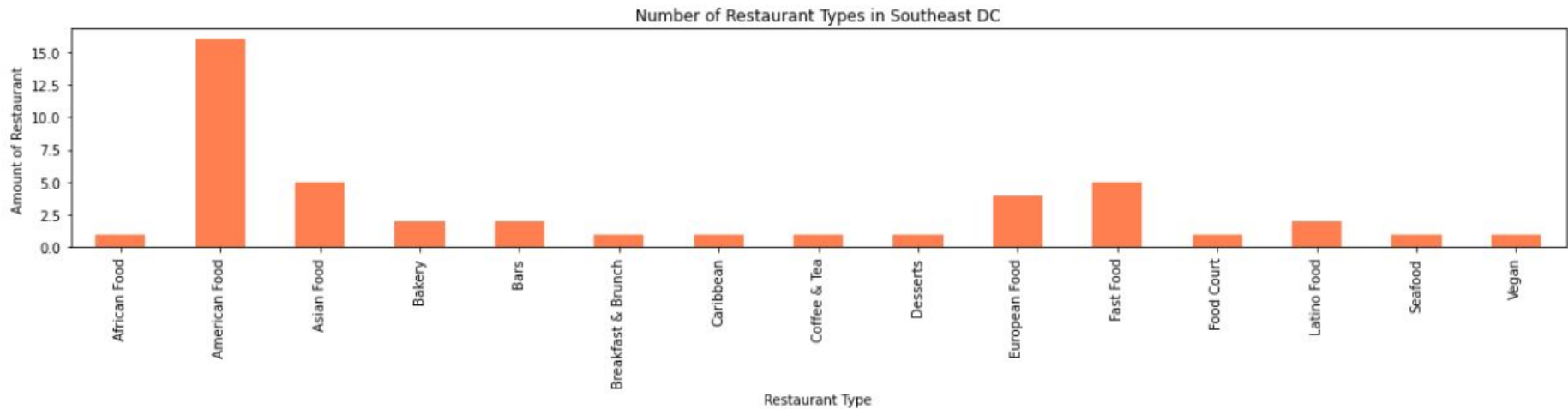
Northeast



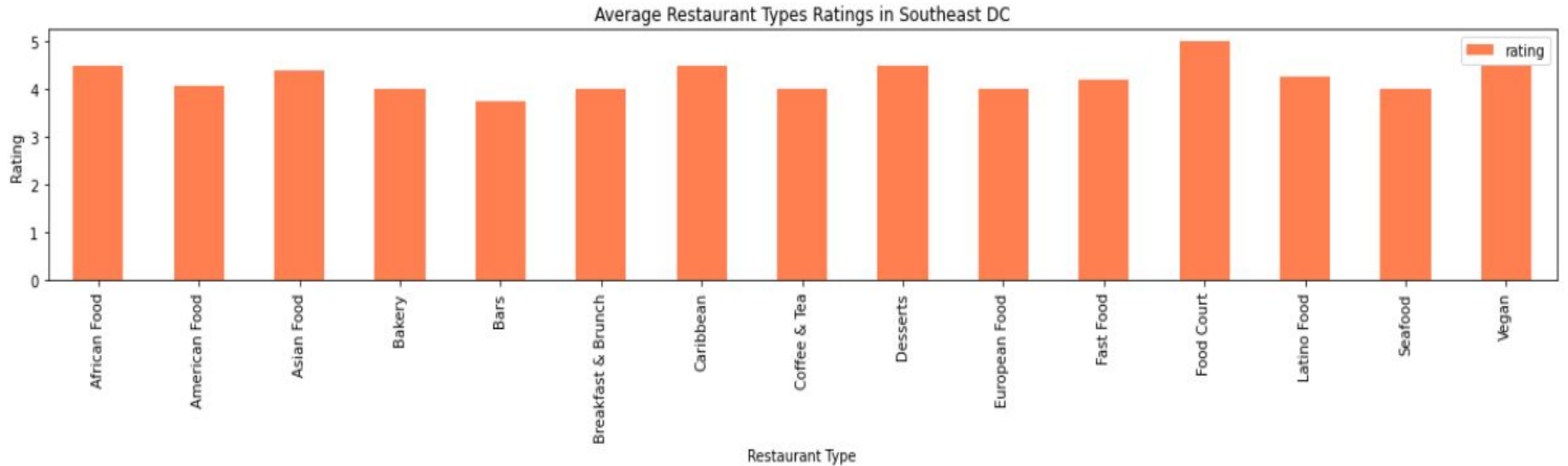
Southeast



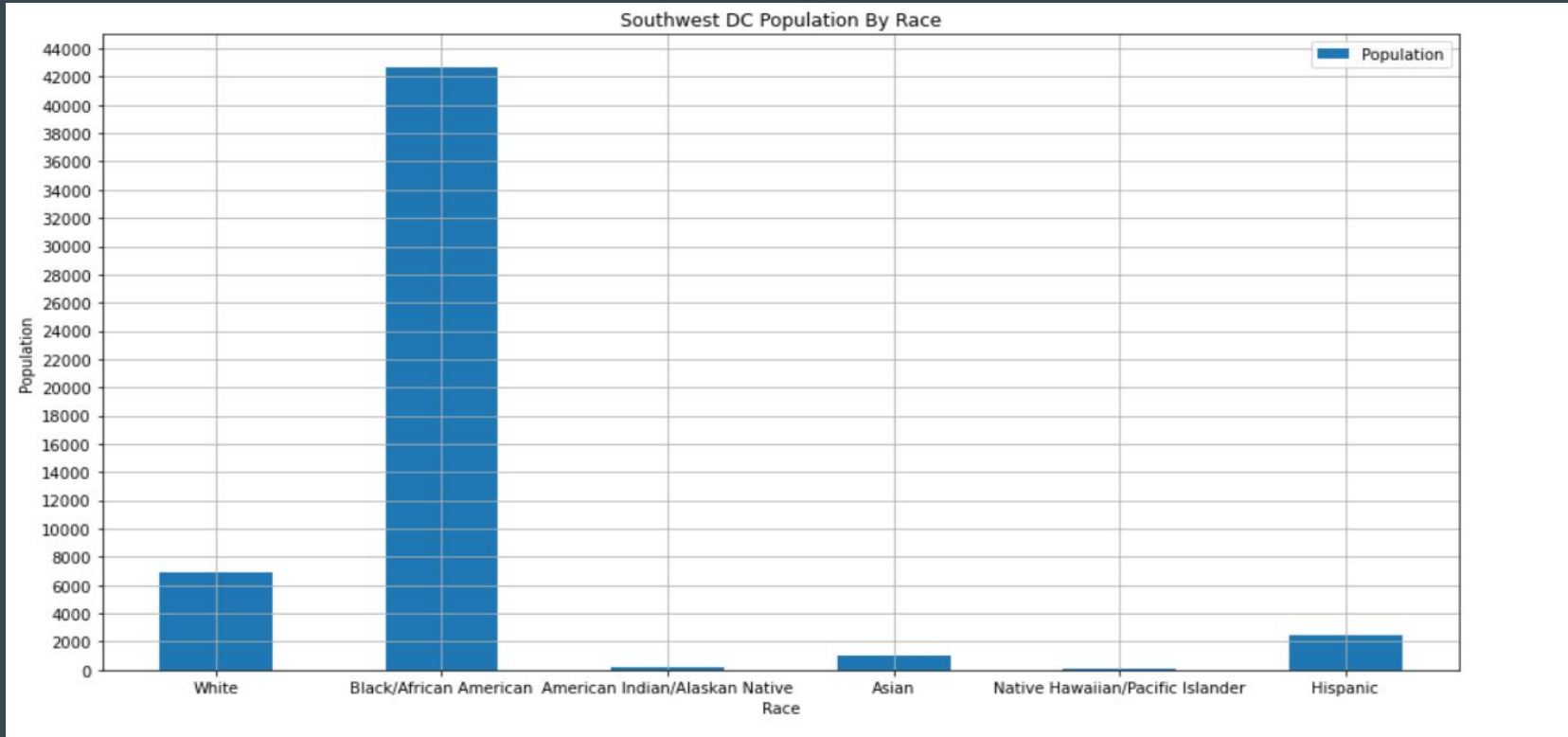
Southeast



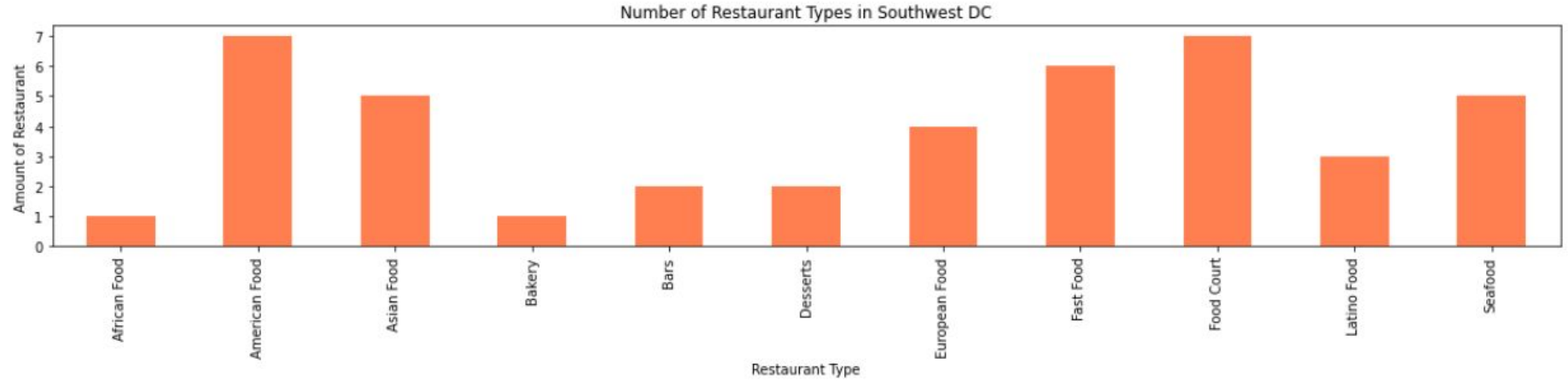
Southeast



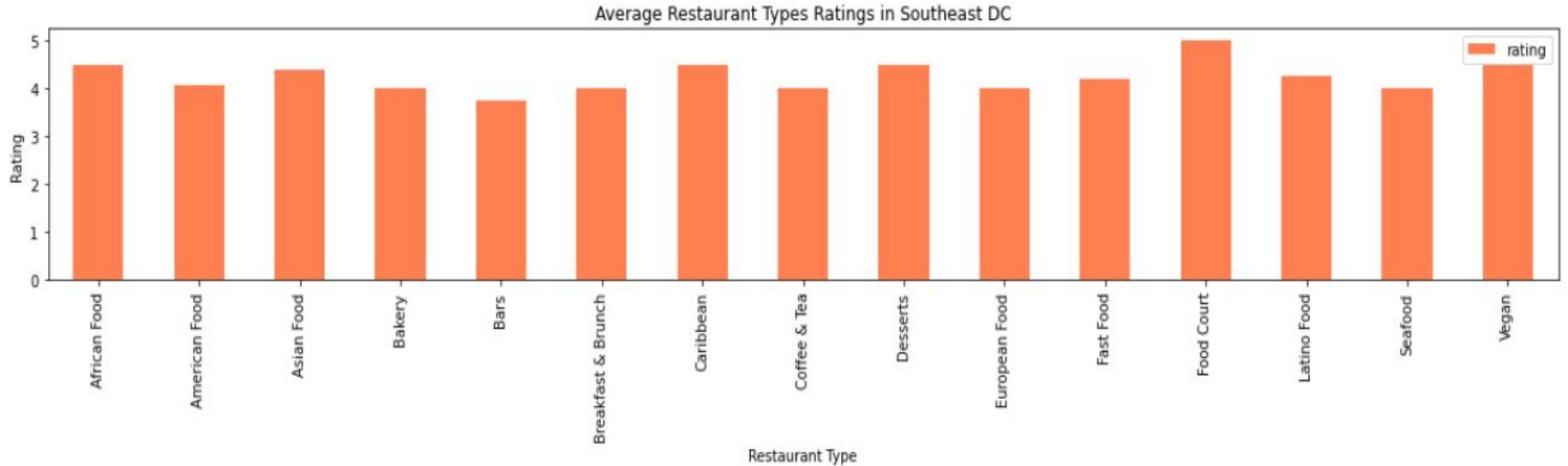
Southwest



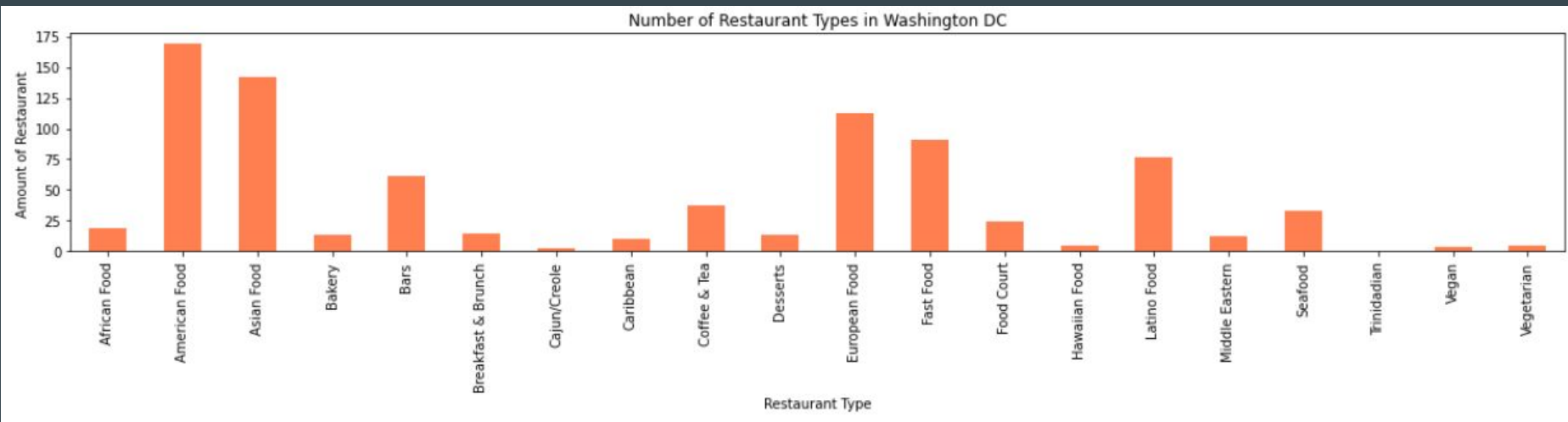
Southwest



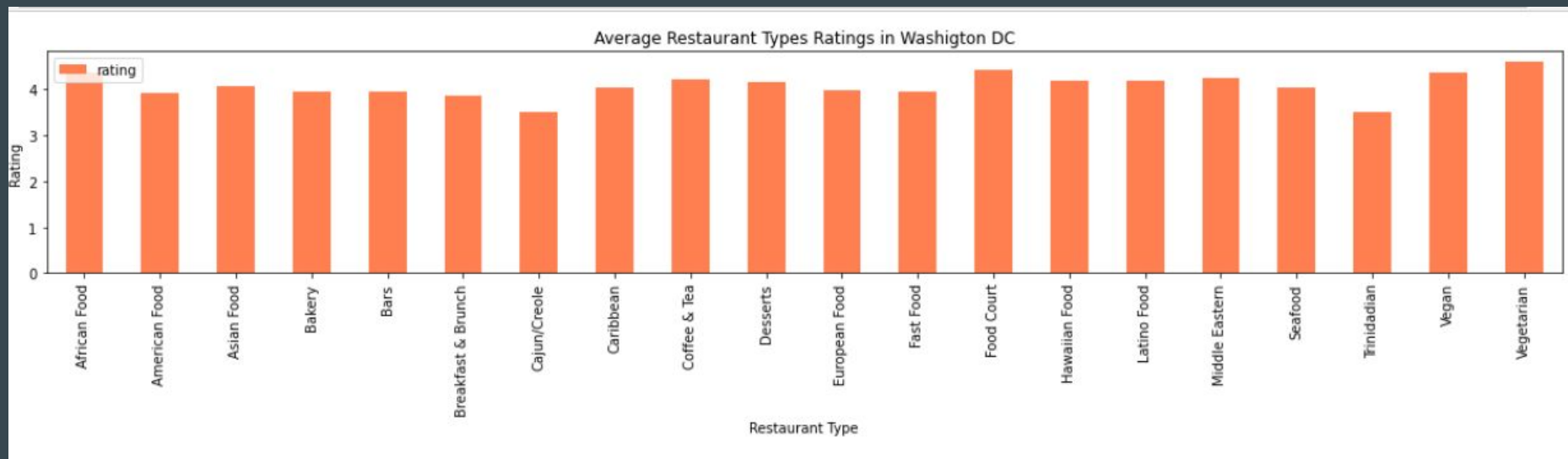
Southwest



Washington DC



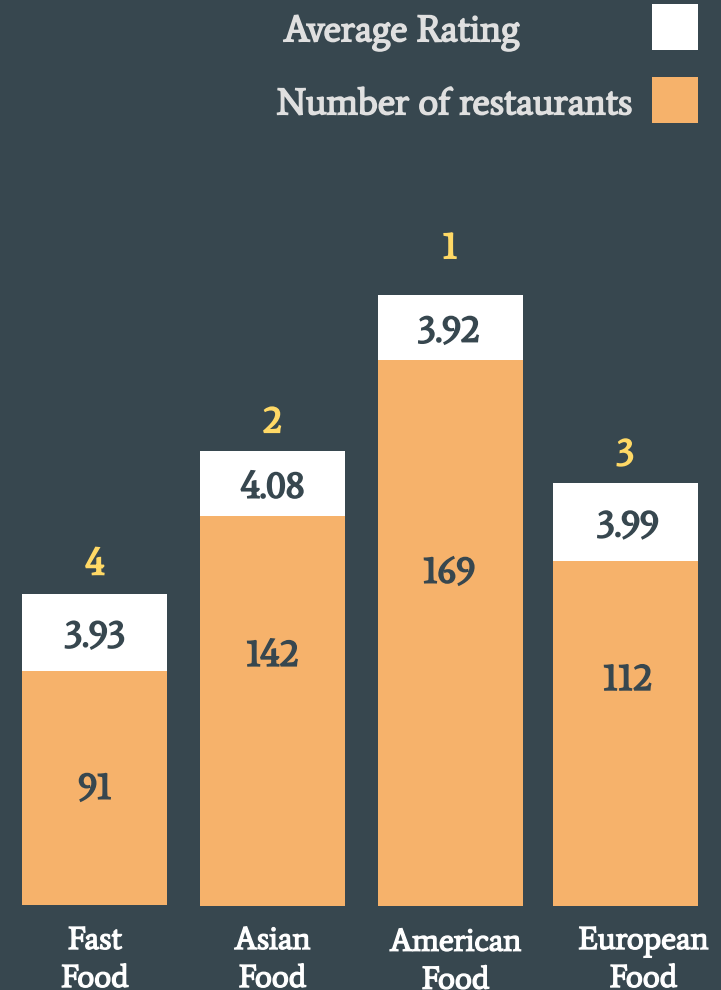
Washington DC



Trend analysis

Findings

- Most competitive restaurant types across each quadrants are American and Asian restaurants.
- Least competitive market for restaurant types are foreign and niche restaurants.
- All restaurant types have an average rating of **3 stars** or more but the restaurant types in Southwest DC have lower average ratings than any other quadrant.
- Despite the competitive nature of Asian restaurants, there are few Asians living in DC compared to White or African American. African Americans are majority in every DC quadrant except in Northwest DC which is majority white. In Northwest DC, there were more european restaurants than any other quadrant.



Food Trucks

- Food trucks do not have a fixed location
- Can sell everywhere and attract tourists !
- Higher average rating than all restaurants with a 4.34 out of 5 !



Insights:

There is more research to be done to build a stronger recommendation.

We didn't anticipate African Americans being the majority in every quadrant of DC except Northwest

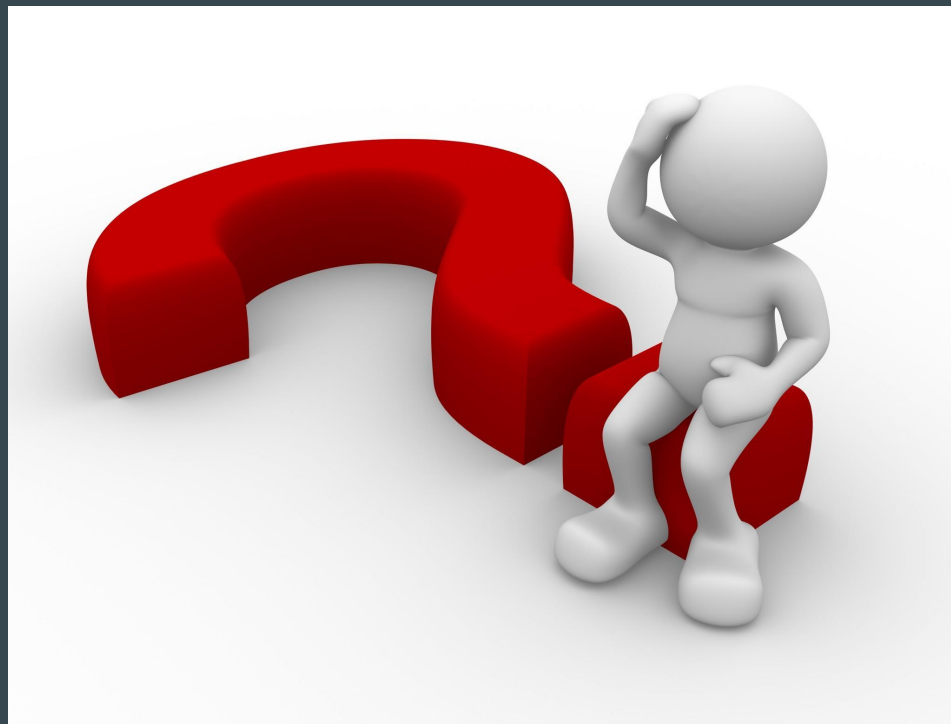
An interesting find is that Asian food is just as competitive as American Restaurants

Another find is that Southeast DC is the only location without a much diversity in restaurant types but yet as the highest amount of African americans.



Problems encountered:

- The Ethnicity Data was only limited to Zip codes and Wards which made it difficult to get a very accurate representation of the four quadrants. Some wards and zip code spanned across multiple quadrants.
- Troubleshooting our code was quite a challenge !



Questions ?

Thank you for listening!

