Write this as a brief summary of your interests and intent, including:

* The kind of data you'd like to work with and the field you're interested in (e.g., trading, quantitative analysis).

|  |  |
| --- | --- |
| [**COVID-19 Data**](https://github.com/M-Media-Group/Covid-19-API) | Get live and historical data regarding Coronavirus cases per country. |

|  |  |
| --- | --- |
| [**VeganCheck**](https://jokenetwork.de/vegancheck-api) | API that provides a couple of information about a food or non-food product based on its EAN/UPC code. |

* The kinds of questions you'll be asking of that data.

Vegan:

* Name of the product
* Is it vegan?
* Is it vegetarian?
* Does it contain palmoil?
* The nutriscore of the product
* If it has been tested on animals
* The VeganCheck Grade
* Where this data comes from

Covid Question

* status
  + Confirmed
  + Deaths
  + Recovered (DEPRECIATED)

Zillow

* Home price housing trend when interest rate went up and down in last 5 years
* We can check the housing trend in the US and cities
* How does the interest rate affect the buying habits of the consumer
* an area undergoing rapid gentrification"
* Possible source for such data.

This constitutes your Project Proposal and Outline, and it should look something like this: Our project is to uncover patterns in credit card fraud. We'll examine relationships between types of transactions and location; purchase prices and times of day; trends in purchases over the course of the year; and related questions, as the data admits.

The technical requirements for Project 1 are as follows.

#1 Calculate the highest sales price in each of the 5 city in a year Manhattan

Bronx

Brooklyn

Queens

State island

#2 Calculate the lowest sales price in each of the 5 city in a year

Manhattan

Bronx

Brooklyn

Queens

State island

#3 Geo-plot the neighborhood housing units sold for each city in a year 2022-2023

Manhattan -

Bronx-

Brooklyn -

Queens -

State island -

#4 Calculate which Building Class category was sold the most for all the 5 neighborhood and build a bar plot to see it for the building classes.

#5 Bar chart – for all 5 cities and each bar sliced by neighborhood

#6 Bar chart - for all 5 cities and each bar sliced by each Building Class category

#7 API – With Zillow [**Zillow Neighborhood Information**](http://www.zillow.com/webtools/neighborhood-data/)

Real estate site Zillow offers APIs that give access to neighborhood information that can be integrated into other applications. (They also offer a number of other APIs, including postings, property details, home valuations, and more.)

A story why are we

import pandas as pd

# Load the dataset into a Pandas DataFrame

df = pd.read\_csv('housing\_sales\_data.csv')

# Filter the data for the year 2022-2023

df = df[(df['Year'] >= 2022) & (df['Year'] <= 2023)]

# Group the data by city

grouped = df.groupby('City')

# Calculate the highest sales price for each city

highest\_prices = grouped['Sale Price'].max()

print("Highest sales prices:")

print(highest\_prices)

# Calculate the lowest sales price for each city

lowest\_prices = grouped['Sale Price'].min()

print("\nLowest sales prices:")

print(lowest\_prices)

import matplotlib.pyplot as plt

# Group the data by city and neighborhood

grouped = df.groupby(['City', 'Neighborhood'])

# Calculate the number of housing units sold in each neighborhood

units\_sold = grouped['Sale Price'].count()

# Plot the results as a bar chart

units\_sold.plot(kind='bar', figsize=(20,10))

plt.xlabel('City and Neighborhood')

plt.ylabel('Number of Housing Units Sold')

plt.title('Neighborhood Housing Units Sold in 2022-2023')

plt.show()

*  Use Pandas to clean and format your dataset(s).
*  Create a Jupyter Notebook describing the **data exploration and cleanup** process.
*  Create a Jupyter Notebook illustrating the **final data analysis**.
*  Use Hvplot or GeoViews to create six to eight data visualizations (ideally, at least two per question asked of the data).
*  Save PNG images of your visualizations to distribute to the class and instructional team and for inclusion in your presentation and your repo's README.md file.
*  Use one new Python library that hasn't been covered in class.
*  Optionally, use at least one API, if you can find an API with data pertinent to your primary research questions.
*  Create a README.md in your repo with a write-up summarizing your major findings. This should include a heading for each question you asked of your data and under each heading a short description of what you found and any relevant plots.

import plotly.express as px  
import pandas as pd# Load the data into a Pandas dataframe  
df = pd.read\_csv('brooklyn\_home\_sale\_prices.csv')# Filter the data to only include sales in Brooklyn  
df = df[df['Borough'] == 'Brooklyn']# Group the data by block and year, and calculate the average sale price for each group  
grouped = df.groupby(['Block', 'Year']).mean()# Create the Choropleth map  
fig = px.choropleth(grouped,  
                    locations='Block',  
                    color='Sale Price',  
                    animation\_frame='Year',  
                    animation\_group='Block',  
                    color\_continuous\_scale='reds',  
                    title='Average Home Sale Price by Block in Brooklyn Over Time')# Show the map  
fig.show()