Project 2 Solutions

Kevin Choe

Collaborators: (Collaborators listed here. Include names, which part of the project you gave or sought help with, and how you helped or were helped.)

TA help: Neha Priyadarshini

Online resources used: (List of links/resources (if any) here. Include web addresses, which part of the project the resource helped with, and how you were helped.)

Question 1

```
import pandas as pd
myDF = pd.read_csv("/class/datamine/data/craigslist/vehicles.csv")
myDF.head()
```

```
           id  ...      long
0  7119256118  ... -114.2690
1  7120880186  ... -123.8240
2  7115048251  ...  -81.9654
3  7119250502  ... -114.2710
4  7120433904  ...  -68.8963

[5 rows x 25 columns]
```

```
from pathlib import Path
p = Path("/class/datamine/data/craigslist/vehicles.csv")
p.stat().st_size
```

```
1235378360
```

```
print(f'my file size is: {p.stat().st_size/1000000}')
```

```
my file size is: 1235.37836
```

Question 2

```
myDF.shape
```

```
(435849, 25)
```

```
print(f'The number of rows are: {myDF.shape[0]}')
```

```
The number of rows are: 435849
```

```
print(f'The number of columns are: {myDF.shape[1]}')
```

```
The number of columns are: 25
```

```
myDF.columns
```

```
Index(['id', 'url', 'region', 'region_url', 'price', 'year', 'manufacturer',
       'model', 'condition', 'cylinders', 'fuel', 'odometer', 'title_status',
       'transmission', 'vin', 'drive', 'size', 'type', 'paint_color',
       'image_url', 'description', 'county', 'state', 'lat', 'long'],
      dtype='object')
```

Question 3

```
columns = list(myDF.columns)
len(columns)
```

25

```
columns.append("extra")
print(f'Second column name is :{columns[1]}')
```

```
Second column name is :url
```

```
print(f'Patterned column names are: {columns[::2]}')
```

```
Patterned column names are: ['id', 'region', 'price', 'manufacturer',
'condition', 'fuel', 'title_status', 'vin', 'size', 'paint_color',
'description', 'state', 'long']
```

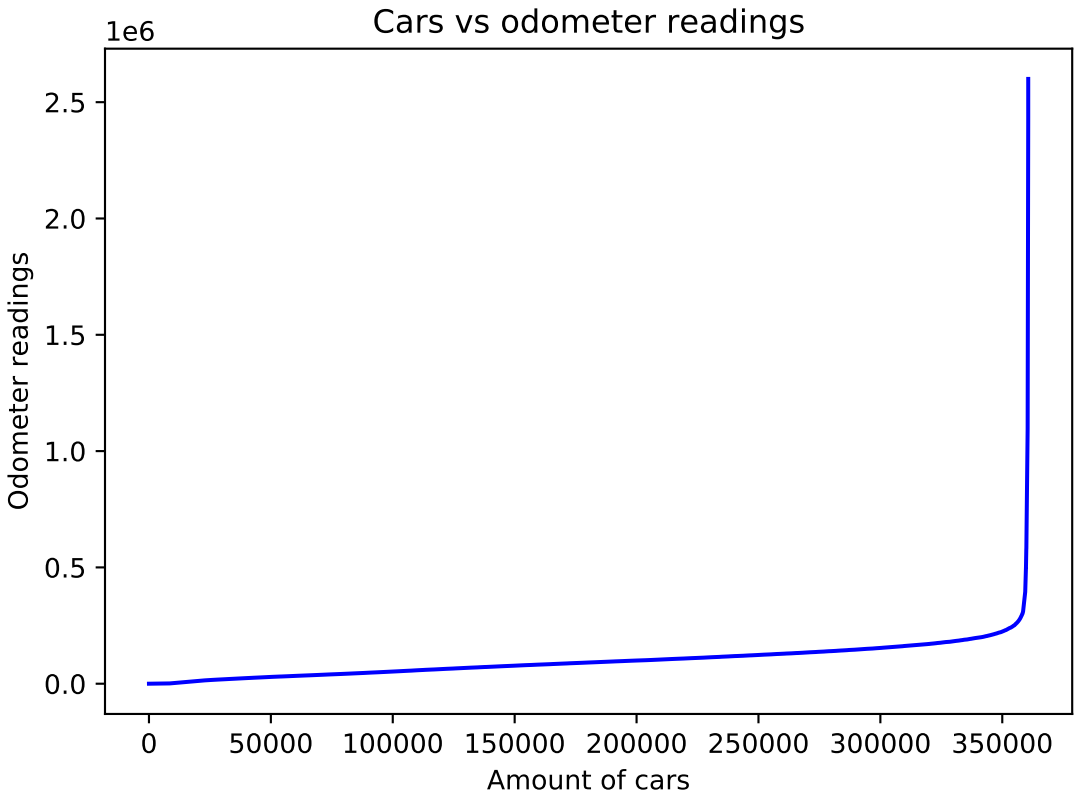
```
print(f'Last four column names are: {columns[-4:]}')
```

```
Last four column names are: ['state', 'lat', 'long', 'extra']
```

```
columns.remove("extra")
```

Question 4

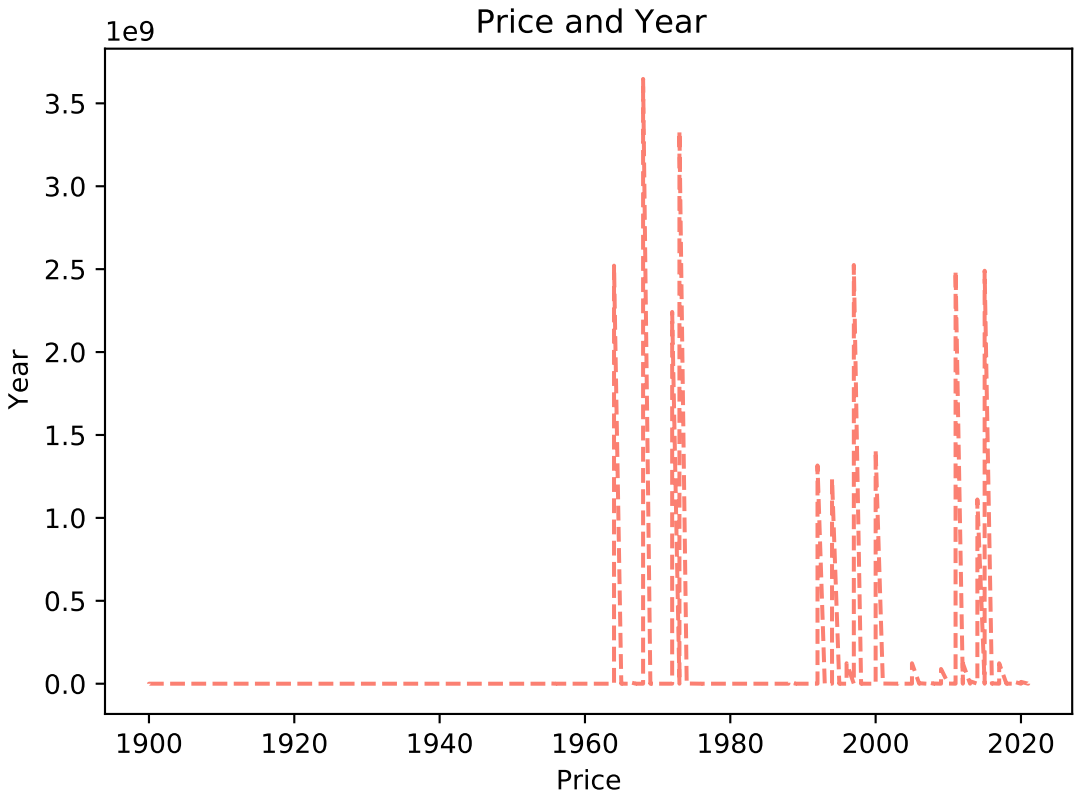
```
from matplotlib import pyplot as plt
my_values = list(tuple(myDF.loc[:, 'odometer'].dropna().to_list()))
my_values.sort()
plt.plot(my_values[0:-50], color="blue")
plt.title("Cars vs odometer readings")
plt.xlabel("Amount of cars")
plt.ylabel("Odometer readings")
plt.show()
```




```
plt.close()
```

Question 5

```
myDF.sort_values(['year', 'price'], inplace = True)
plt.plot(myDF['year'], myDF['price'], color="#FB8072", linestyle="dashed")
plt.title("Price and Year")
plt.xlabel("Price")
plt.ylabel("Year")
plt.show()
```




```
plt.close()
```

Pledge

By submitting this work I hereby pledge that this is my own, personal work. I've acknowledged in the designated place at the top of this file all sources that I used to complete said work, including but not limited to: online resources, books, and electronic communications. I've noted all collaboration with fellow students and/or TA's. I did not copy or plagiarize another's work.

> As a Boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together – We are Purdue.