Programming Assignment 3

Name: Jimmy Nguyen Date: October 31, 2022 Course/Professor: INST447/Sigalo

---

Research Question: Is there an association between average or median household income and the number of Parking Enforcement service requests?

IMPORT FUNCTIONS

In [9]:
```
import matplotlib.pyplot as plt
%matplotlib inline
import pandas as pd
```

Read in the two files (WardIcome.csv & _311requests.csv)

In [10]:
```
req_df = pd.read_csv("_311requests.csv")
ward_df = pd.read_csv("WardIncome.csv")
```

Subset the _311requests data

In [11]:
```
parkenf = req_df[['SERVICECODEDESCRIPTION', 'WARD']].groupby(['SERVICECODEDESCRIPTION'
parkenf[72:73]
```

Out[11]:

	WARD
SERVICECODEDESCRIPTION	
Parking Enforcement	26714

In [12]:
```
order = req_df.sort_values('SERVICECODEDESCRIPTION')
order = order[['SERVICECODEDESCRIPTION', 'WARD']]
order = order[110832:137546]
order
```

Out[12]:

	SERVICECODEDESCRIPTION	WARD
312398	Parking Enforcement	Ward 2
279776	Parking Enforcement	Ward 8
200012	Parking Enforcement	Ward 7
79460	Parking Enforcement	Ward 6
61006	Parking Enforcement	Ward 2
...	...	...
65172	Parking Enforcement	Ward 2
274519	Parking Enforcement	Ward 1
99199	Parking Enforcement	Ward 2
173095	Parking Enforcement	Ward 3
269143	Parking Enforcement	Ward 6

26714 rows × 2 columns

To explain the process of subsetting, I first decided to find the count of all the parking enforcements within the SERVICECODEDESCRIPTION column. The count came out to be '26714'. After, I sorted out the values within 'SERVICECODEDESCRIPTION' by 'ALPHABETICAL' order. I spent some time finding the first entry of the 'Parking Enforcement' and added the additional '26714' to get the end entry of 'Parking Enforcement'

Aggregation of subsetted 311 requests dataset

In [17]:
```
parkenf = order[['WARD','SERVICECODEDESCRIPTION']].groupby(['WARD']).agg('count')
parkenf = parkenf.reset_index()
parkenf.columns = ['Ward', 'Parking Enforcements Per Ward']
parkenf
```

Out[17]:

	Ward	Parking Enforcements Per Ward
0	Ward 1	4282
1	Ward 2	5572
2	Ward 3	2139
3	Ward 4	2048
4	Ward 5	3599
5	Ward 6	6318
6	Ward 7	1405
7	Ward 8	1351

The median income, mean income, and number of parking enforcement requests based on merging

In [18]:
```
merge = pd.merge(ward_df, parkenf, how = 'inner', on = ['Ward'])
merge
```

Out[18]:

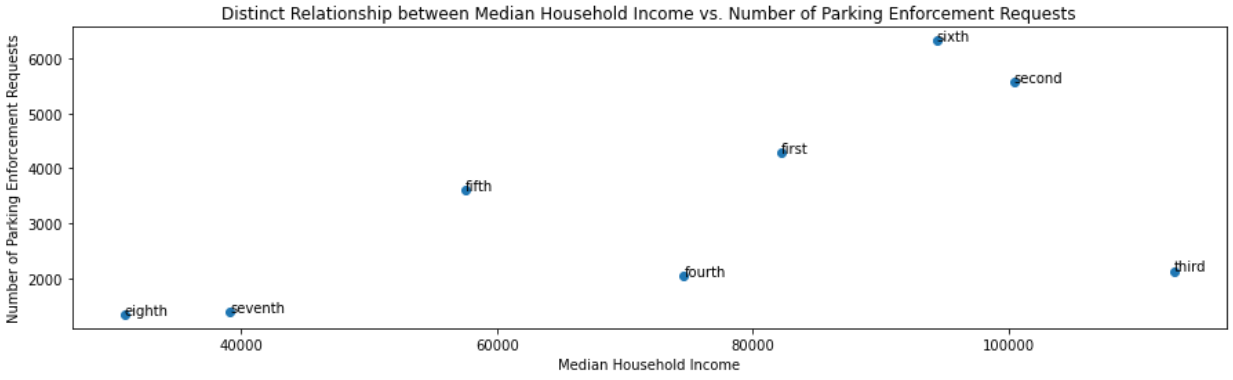
	Ward	MedianHHincome	MeanHHincome	Parking Enforcements Per Ward
0	Ward 1	82159	106388	4282
1	Ward 2	100388	140459	5572
2	Ward 3	112873	176921	2139
3	Ward 4	74600	111933	2048
4	Ward 5	57554	77329	3599
5	Ward 6	94343	117002	6318
6	Ward 7	39165	53093	1405
7	Ward 8	30910	44245	1351

A graph/plot to find the relationship between median household income and the number of parking enforcement requests

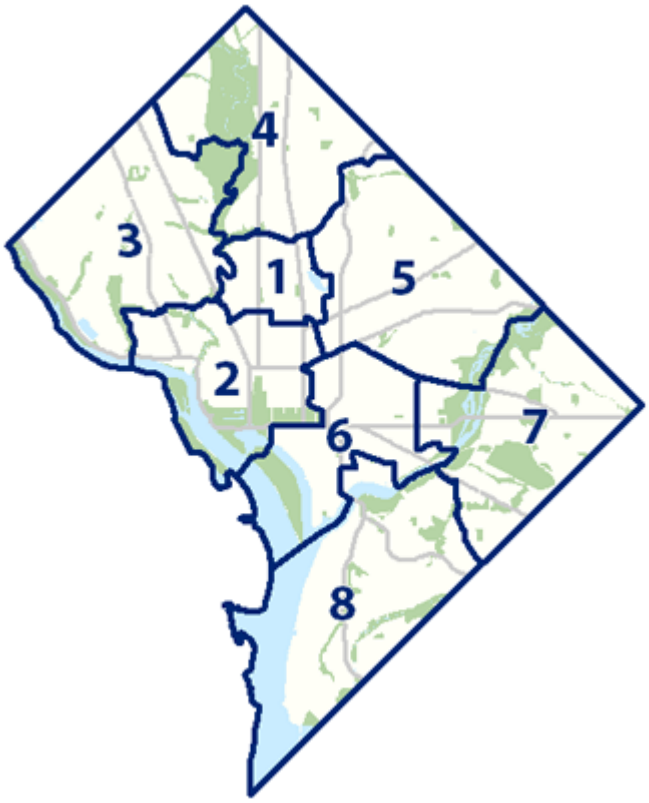
In [21]:
```
fig = plt.figure(figsize = (15, 4))
x = merge['MedianHHincome']
y = merge['Parking Enforcements Per Ward']
text = ["first", "second", "third", "fourth", "fifth", "sixth", "seventh", "eighth"]
plt.scatter(x, y)

for i in range(len(x)):
    plt.annotate(text[i], (x[i], y[i]))

plt.xlabel('Median Household Income')
plt.ylabel('Number of Parking Enforcement Requests')
plt.title('Distinct Relationship between Median Household Income vs. Number of Parking
plt.show()
```



Graph 1 Interpretation



There are a few noticable key points that we can make comparisons about when observing the district wards with the dataset above. One of the main key points noted during this analysis is that 'Wards 6, 2, 1' have the high number of parking requests. These regions are located in the deep inner sections of DC, which suggests that DC police are highly concerned about regions where there are lots of people on a daily average. Another key point which ties into the first one is that the regions outside of inner DC (Wards 8, 7, 5, 4, 3) are less suseptible to parking enforcements requests. This also makes another point as people are not actively requests for law enforcements to check these regions due to less traffic. The last key point that I'd like to make is that Ward 5 is on the outer region of DC but also has a great amount of parking enforcement requests. My theory is that this section of DC could become a popular section for big events which would pull in more traffic, making it more susceptible to high volumes of parking enforcement request calls.

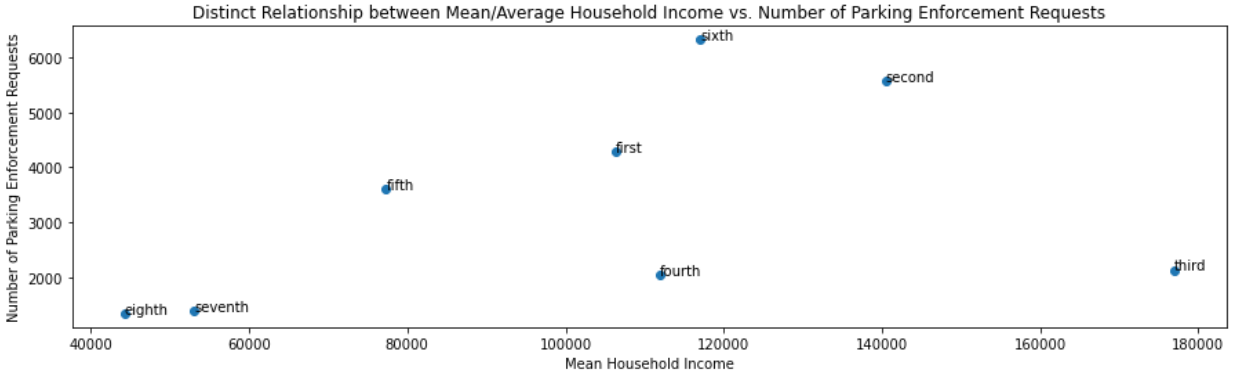
From this analysis, I have determined that the 'median household income' DOES NOT have any corellation with the 'number of parking enforcement requests'. More so, I have determined that the 'number of parking enforcement requests' are based on high traffic locations within Washington DC.

Create a graph/plot to find the relationship between mean household income and the number of parking enforcement requests

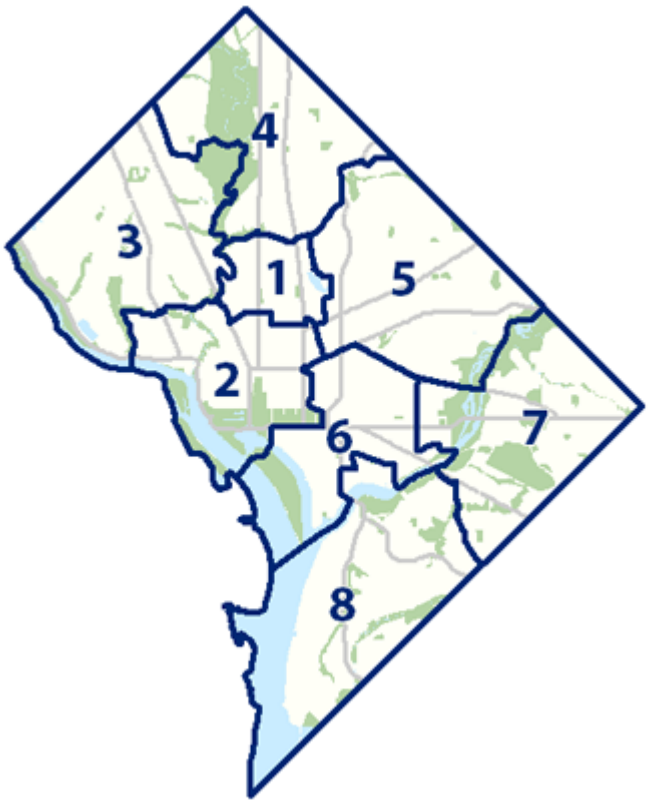
In [22]:
```
fig = plt.figure(figsize = (15, 4))
x = merge['MeanHHincome']
y = merge['Parking Enforcements Per Ward']
text = ["first", "second", "third", "fourth", "fifth", "sixth", "seventh", "eighth"]
plt.scatter(x, y)

for i in range(len(x)):
    plt.annotate(text[i], (x[i], y[i]))
```

```
plt.xlabel('Mean Household Income')
plt.ylabel('Number of Parking Enforcement Requests')
plt.title('Distinct Relationship between Mean/Average Household Income vs. Number of P
plt.show()
```



Graph 2 Interpretation



Once again, there is a noticable increase in Wards 1, 2, and 6. These wards are located deeper within DC and are susceptible to parking enforcement requests due to their high traffic. A major difference when we compare the two graphs is that the mean/average of household incomes spans much more greatly. To further explain, the range of the median household income spans from 40,000 to around 120,000 while the the range of the mean/average household income spans from 40,000 towards up to 18,000+. Another distinct difference when we further analyze the mean and median of each ward is that Ward 4 is a comparitively less of a median household income than to Ward 1, but has a greater mean/average household income than to Ward 1, which I found was very interesting.

From this analysis, I have determined that the 'mean household income' DOES NOT have any corellation with the 'number of parking enforcement requests'.

In [ ]: