hw_1-Copy1

November 9, 2022

1 Homework 1

1.1 In-class exercise

1.1.1 Step 1

Read in the 311 dataset from:

https://storage.googleapis.com/python-public-policy/data/311_requests_2018-19_sample.csv.zip

and save it as `df`. Remember there is a package you need to import in order for this to work.

[2]:
```
import pandas as pd

df = pd.read_csv(
    "https://storage.googleapis.com/python-public-policy/data/
 ↪311_requests_2018-19_sample.csv.zip"
)
```

```
/tmp/ipykernel_177/14146379.py:3: DtypeWarning: Columns (8,20,31,34) have mixed
types. Specify dtype option on import or set low_memory=False.
  df = pd.read_csv(
```

1.1.2 Step 2

Answer these two questions about the data, showing your code that produced the result in two individual code cells:

1. What is the minimum value in the `Created Date` column? Try using `.min()`.
2. What is the maximum value in the `Created Date` column? Try using `.max()`.

[4]:
```
df["Created Date"].min()
```

[4]:
```
'01/01/2019 01:00:00 PM'
```

[9]:
```
df["Created Date"].max()
```

[9]:
```
'12/31/2018 12:59:37 AM'
```

1.1.3 Step 3

Does anything about the resulting min and max values surprise you? What do think causes this?

Hint: Look at the year. If you use `.head()` and `.tail()` you can find the real min and max dates. We'll learn more about how to properly handle dates in a later lecture.

YOUR RESPONSE HERE

1.2 Coding: Exploring Trends in How People Submit 311 Requests

Use this zipped copy of the `311_requests_2018-19_sample_clean.csv` created during our lecture to complete the homework. Read it in from:

https://storage.googleapis.com/python-public-policy/data/311_requests_2018-19_sample_clean.csv.zip

[11]:
```
df["Created Date"].head()
```

[11]:
```
0    08/01/2018 12:05:13 AM
1    08/01/2018 12:06:05 AM
2    08/01/2018 12:06:16 AM
3    08/01/2018 12:06:29 AM
4    08/01/2018 12:06:51 AM
Name: Created Date, dtype: object
```

[12]:
```
df["Created Date"].tail()
```

[12]:
```
499995    08/24/2019 01:46:09 AM
499996    08/24/2019 01:49:49 AM
499997    08/24/2019 01:56:35 AM
499998    08/24/2019 01:56:40 AM
499999    08/24/2019 01:57:58 AM
Name: Created Date, dtype: object
```

1.2.1 Q1

Which complaint submission method (`Open Data Channel Type`) is used most frequently, according to this dataset? Provide the count of each for comparison.

[7]:
```
display(df.groupby('Open Data Channel Type').size())
```

```
Open Data Channel Type
MOBILE      64566
ONLINE     102898
OTHER        3502
PHONE      256558
UNKNOWN     72476
dtype: int64
```

1.2.2 Step 2

Get the count of each submission method per complaint type, sorted by complaint type.

[13]:
```
count_submissionmethod = df. groupby(['Complaint Type','Open Data Channel␣
↪Type']).size().reset_index(name="count").sort_values('count',␣
↪ascending=False).head(10)
count_submissionmethod
```

[13]:

	Complaint Type	Open Data Channel Type	count
344	Request Large Bulky Item Collection	PHONE	30939
178	HEAT/HOT WATER	PHONE	22604
287	Noise - Residential	PHONE	16810
46	Blocked Driveway	PHONE	16692
286	Noise - Residential	ONLINE	15549
213	Illegal Parking	MOBILE	13515
395	Street Light Condition	UNKNOWN	13425
435	UNSANITARY CONDITION	PHONE	11066
215	Illegal Parking	PHONE	11064
394	Street Condition	UNKNOWN	10679

Review the results and share a sentence or two with your observations.

The maximum number of complaint is about requesting large bulky item collection and it is made by phone.

[ ]:
```
##Step 3
Provide a simplified table that only lists the most common request submission␣
↪method for each complaint type. There should be one row per complaint type,␣
↪capturing the most common submission method for that complaint type.
```

[41]:
```
sorted_complaintcounts = count_submissionmethod.
↪sort_values('count',ascending=False)
sorted_complaintcounts
max_complaintmethod = sorted_complaintcounts
max_complaintmethod
```

[41]:

	Complaint Type	Open Data Channel Type	count
344	Request Large Bulky Item Collection	PHONE	30939
178	HEAT/HOT WATER	PHONE	22604
287	Noise - Residential	PHONE	16810
46	Blocked Driveway	PHONE	16692
286	Noise - Residential	ONLINE	15549
213	Illegal Parking	MOBILE	13515
395	Street Light Condition	UNKNOWN	13425
435	UNSANITARY CONDITION	PHONE	11066
215	Illegal Parking	PHONE	11064
394	Street Condition	UNKNOWN	10679

Can you provide a potential explaination for why some complaints are more frequently made via phone and mobile versus on the website?

Many complaint types requires an urgent and direct response where calling by phone is the easiest, such as Heat/Hot Water, Noise Residential etc. However, website complaints require longer time to process, and people like the elderly may not prefer to use website.

Did you work with any other students on this assignment?

Yes, I talked with a friend who is not in our class.

Now turn in the assignment.

1.3 Tutorials

1. Glance through pandas' comparison with other tools for any you are familiar with
2. Filtering/indexing `DataFrame`s
   - Filter specific rows from a `DataFrame`
   - Boolean indexing
3. Video introducing functions
4. Coding Style Guides - Please skim these; I don't expect you to understand and follow everything in them. The most important guidelines to pay attention to are indentation and keeping each statement on its own line.
   - The Hitchhiker's Guide to Python
   - PEP 8
5. Guide to commenting your code
6. Quartz Guide to Bad Data

1.3.1 Optional

More on indexing:

- How to Select Rows from Pandas DataFrame
- Selecting Subsets of Data in Pandas: Part 1 and Part 2

[ ]: