

In [4]: `import pandas as pd`

In [8]: `# 1. Create a DataFrame for the 201908-citibike-tripdata data.  
df = pd.read_csv("201908-citibike-tripdata/201908-citibike-tripdata.csv")  
df.head()`

Out[8]:

	tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude	end station id	end station name
0	393	2019-08-01 00:00:01.4680	2019-08-01 00:06:35.3780	531.0	Forsyth St & Broome St	40.718939	-73.992663	408.0	Market St & Chrystie St
1	627	2019-08-01 00:00:01.9290	2019-08-01 00:10:29.7840	274.0	Lafayette Ave & Fort Greene Pl	40.686919	-73.976682	3409.0	Bergen St & Smith St
2	1132	2019-08-01 00:00:04.0480	2019-08-01 00:18:56.1650	2000.0	Front St & Washington St	40.702551	-73.989402	3388.0	Presidential St & Henry St
3	1780	2019-08-01 00:00:04.1630	2019-08-01 00:29:44.7940	479.0	9 Ave & W 45 St	40.760193	-73.991255	473.0	Rivington St & Chrystie St
4	1517	2019-08-01 00:00:05.4580	2019-08-01 00:25:23.4550	3312.0	1 Ave & E 94 St	40.781721	-73.945940	3312.0	1 Ave & E 94 St

In [7]: `# 2. Check the datatypes of your columns.  
df.dtypes`

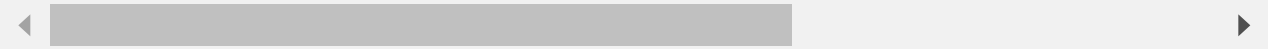
Out[7]:

tripduration	int64
starttime	object
stoptime	object
start station id	float64
start station name	object
start station latitude	float64
start station longitude	float64
end station id	float64
end station name	object
end station latitude	float64
end station longitude	float64
bikeid	int64
usertype	object
birth year	int64
gender	int64
dtype:	object

In [9]: `# 3. Convert the 'tripduration' column to datetime datatype.  
df.tripduration = pd.to_datetime(df.tripduration, unit = "s")  
df.head()`

Out[9]:

	tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude	end station id	end station name
0	1970-01-01 00:06:33	2019-08-01 00:00:01.4680	2019-08-01 00:06:35.3780	531.0	Forsyth St & Broome St	40.718939	-73.992663	408.0	Market St & Chrystie St
1	1970-01-01 00:10:27	2019-08-01 00:00:01.9290	2019-08-01 00:10:29.7840	274.0	Lafayette Ave & Fort Greene Pl	40.686919	-73.976682	3409.0	Bergen St & Smith St
2	1970-01-01 00:18:52	2019-08-01 00:00:04.0480	2019-08-01 00:18:56.1650	2000.0	Front St & Washington St	40.702551	-73.989402	3388.0	Presidential St & Henry St
3	1970-01-01 00:29:40	2019-08-01 00:00:04.1630	2019-08-01 00:29:44.7940	479.0	9 Ave & W 45 St	40.760193	-73.991255	473.0	Rivington St & Chrystie St
4	1970-01-01 00:25:17	2019-08-01 00:00:05.4580	2019-08-01 00:25:23.4550	3312.0	1 Ave & E 94 St	40.781721	-73.945940	3312.0	1 Ave & E 94 St



In [10]: `# 4. Check the datatypes of your columns.  
df.dtypes`

Out[10]:

tripduration	datetime64[ns]
starttime	object
stoptime	object
start station id	float64
start station name	object
start station latitude	float64
start station longitude	float64
end station id	float64
end station name	object
end station latitude	float64
end station longitude	float64
bikeid	int64
usertype	object
birth year	int64
gender	int64
dtype:	object

In [11]: `# 5. Export the Dataframe as a new CSV file without the index.  
df.to_csv("cleancitibike.csv",index = False)`

In [ ]: