Data Analysis assignment (20 points)

You have been given the dataset `travel-times` in a CSV format. This dataset comes from a driver that uses an app to track GPS coordinates as he drives to work and back each day. The app collects the location and elevation data. In total, data for about 200 trips are summarized in this data set.

Load the `travel-times` in a `df` variable using `pandas` and then perform the following:

- print the shape of the dataset **(1 points)**
- print the first 15 rows of the dataset **(1 points)**
- get information for the features (columns) with missing values **(1 points)**
- drop duplicate values (if any) by keeping only the last instance **(1 points)**
- calculate the total number of missing values (if any) on each column **(2 points)**
- create two copies of the dataframe, and then:
  - drop rows with missing values from the 1st copy **(1 points)**
  - drop columns with missing values from the 2nd copy **(1 points)**
- get summary statistics and see the correlation between the numerical columns **(1 points)**
- show rows 11 to 14 **(1 points)**
- create a subset with trips occurred on November 23, 2011 and January 6, 2012 **(2 points)**
- produce a scatterplot between `Distance` and `TotalTime` **(1 points)**
  - Use:

```
import matplotlib.pyplot as plt
plt.rcParams.update({'font.size': 20, 'figure.figsize': (10, 8)})
```

- produce boxplots for `AvgSpeed` and `AvgMovingSpeed` (use different cells for each) **(2 points)**

```
import matplotlib.pyplot as plt
import pandas as pd

plt.rcParams.update({'font.size': 20, 'figure.figsize': (10, 8)})
```

```
import pandas as pd
```

```
path="/travel-times (1).csv"
df=pd.read_csv(path)
df.head(15)
```

	Date	StartTime	DayOfWeek	GoingTo	Distance	MaxSpeed	AvgSpeed	AvgMovingSpe
0	1/6/2012	16:37	Friday	Home	51.29	127.4	78.3	8
1	1/6/2012	08:20	Friday	GSK	51.63	130.3	81.8	8
2	1/4/2012	16:17	Wednesday	Home	51.27	127.4	82.0	8
3	1/4/2012	07:53	Wednesday	GSK	49.17	132.3	74.2	8
4	1/3/2012	18:57	Tuesday	Home	51.15	136.2	83.4	8
5	1/3/2012	07:57	Tuesday	GSK	51.80	135.8	84.5	8
6	1/2/2012	17:31	Monday	Home	51.37	123.2	82.9	8
7	1/2/2012	07:34	Monday	GSK	49.01	128.3	77.5	8
8	12/23/2011	08:01	Friday	GSK	52.91	130.3	80.9	8
9	12/22/2011	17:19	Thursday	Home	51.17	122.3	70.6	7
10	12/22/2011	08:16	Thursday	GSK	49.15	129.4	74.0	8
11	12/21/2011	07:45	Wednesday	GSK	51.77	124.8	71.7	7
12	12/20/2011	16:05	Tuesday	Home	51.45	130.1	75.2	8
13	12/20/2011	06:04	Tuesday	GSK	49.01	119.0	77.4	8
14	12/19/2011	16:18	Monday	Home	51.04	132.2	77.5	8

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
df.drop_duplicates(inplace=True)
```

```
df.isnull().sum()
```

```
Date                0
StartTime           0
DayOfWeek           0
GoingTo             0
Distance            0
MaxSpeed            0
AvgSpeed            0
AvgMovingSpeed      0
FuelEconomy        17
TotalTime           0
MovingTime          0
Take407All          0
Comments          181
dtype: int64
```

```
df.select_dtypes(exclude='object').corr()
```

	Distance	MaxSpeed	AvgSpeed	AvgMovingSpeed	TotalTime	MovingTime
Distance	1.000000	0.145091	-0.006445	0.011874	0.197207	0.197044
MaxSpeed	0.145091	1.000000	0.253869	0.257823	-0.198775	-0.222574
AvgSpeed	-0.006445	0.253869	1.000000	0.872143	-0.877806	-0.835814
AvgMovingSpeed	0.011874	0.257823	0.872143	1.000000	-0.856986	-0.944433
TotalTime	0.197207	-0.198775	-0.877806	-0.856986	1.000000	0.920935
MovingTime	0.197044	-0.222574	-0.835814	-0.944433	0.920935	1.000000

```
df.iloc[10:14]
```

```
df[df['Date'] == '11/23/2011']
```

	Date	StartTime	DayOfWeek	GoingTo	Distance	MaxSpeed	AvgSpeed	AvgMovingSpeed	FuelEconomy	TotalTime	MovingTime	Take407A
44	11/23/2011	16:17	Wednesday	Home	60.32	129.4	68.9	74.6	9.3	52.5	48.5	N
45	11/23/2011	07:22	Wednesday	GSK	51.60	126.4	67.3	73.6	9.3	46.0	42.1	N

```
df[df['Date'] == '01/06/2012']
```

Date	StartTime	DayOfWeek	GoingTo	Distance	MaxSpeed	AvgSpeed	AvgMovingSpeed	FuelEconomy	TotalTime	MovingTime	Take407All	Comm

```
df.shape
```

```
df.head(15)
```

```
df.info()
```

```
df.plot(kind='scatter', x='Distance', y='TotalTime')
```

```
df('AvgMovingSpeed').plot(kind='box')
```

```
df('AvgMovingSpeed', 'AvgSpeed').plot(kind='box')
```

```
df('AvgSpeed').plot(kind='box')
```