

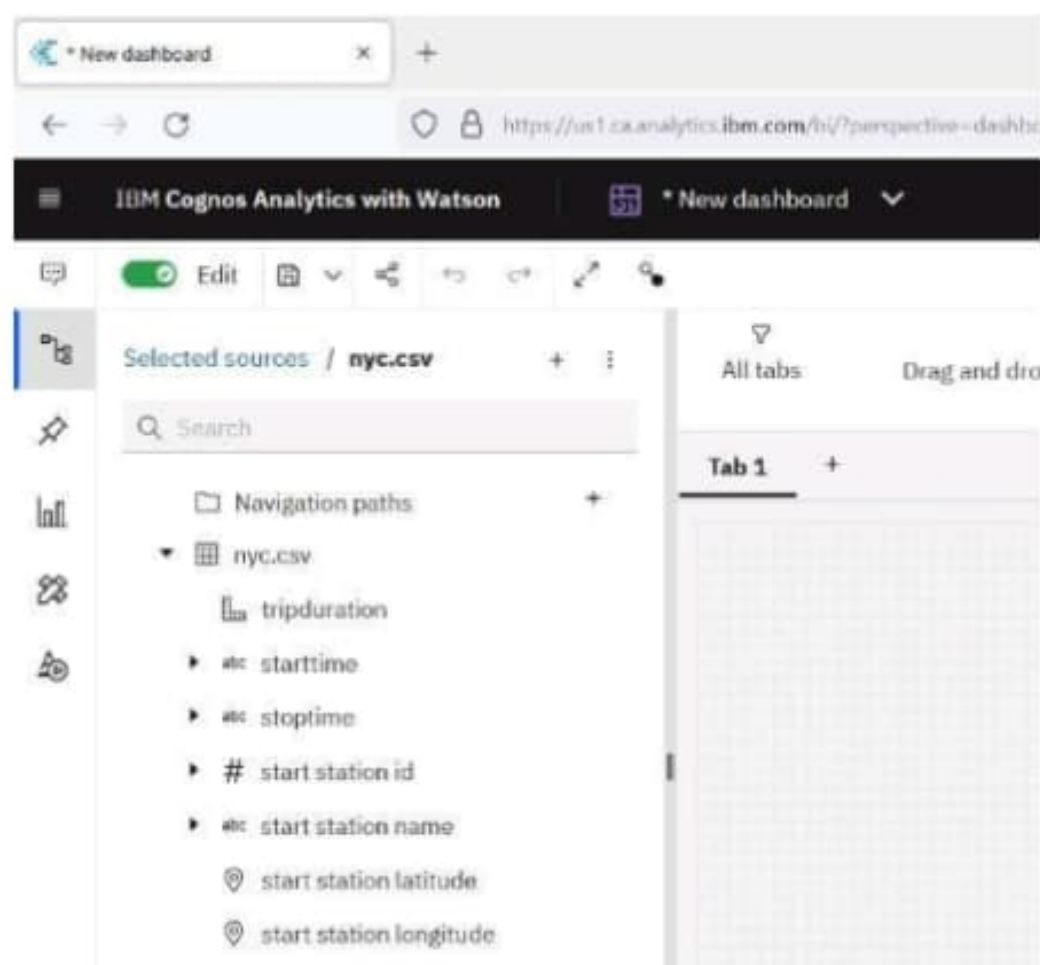
## Project Development Phase – Sprint1

Date	24 October 2022
Team ID	PNT2022TMID45599
Project Name	A new hint to transportation – Analysis of the NYC bike share system.
Maximum Marks	20 Marks

# Data Preparation:

Dataset link : <https://s3.amazonaws.com/tripdata/index.html>

## Uploading the dataset



## Data Cleaning: Finding the

duplicates: `df` =

`data[data.birthyear.notnull()]`

`df.duplicated().sum()`



```
birthyear
1980.0
1969.0
1975.0
1984.0
1994.0
```

0 # There are no null values present in birthyear

Checking null values:

```
d=df.isna().sum()
```

```
d.head()
```

```
[>] tripduration      0
     starttime        0
     stoptime         0
     start station id  47
     start station name 47
     dtype: int64
```

Replace null values:

```
df.replace(np.nan,'-',inplace =
```

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

```
tripduration      0
starttime         0
stoptime          0
start station id  0
start station name 0
start station latitude 0
start station longitude 0
end station id    0
end station name  0
end station latitude 0
end station longitude 0
bikeid           0
usertype          0
birthyear         0
gender            0
tripduration_bins 0
Age              0
dtype: int64
```