PyBer Analysis

DataFrame Tree

Data Exploratory Phase

city_data.cvs

ride_data.cvs

city_data_to_load

ride_data_to_load

city_data_df

ride_data_df

Merge on "city" and create pyber_data_df
(city, data, fare, ride_id, driver_count, type)

Operations Performed on df's
- Count
- Isnull.sum
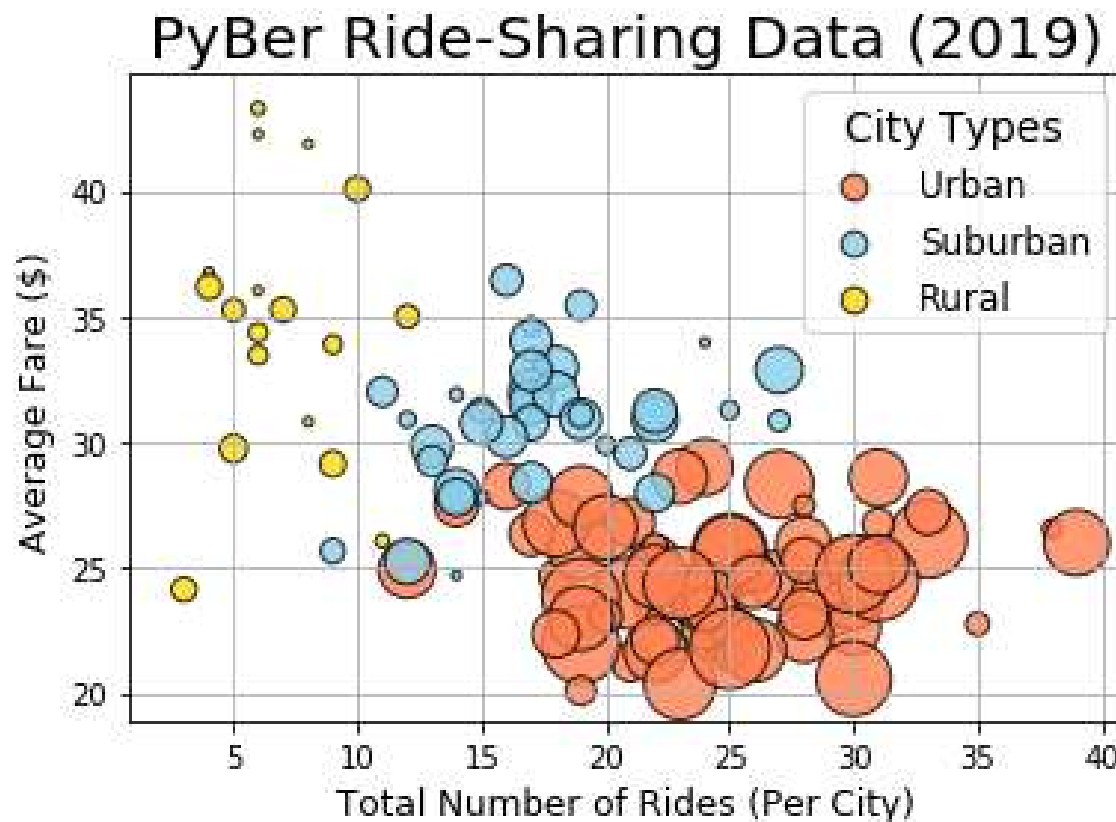- Dtypes
- Type.unique
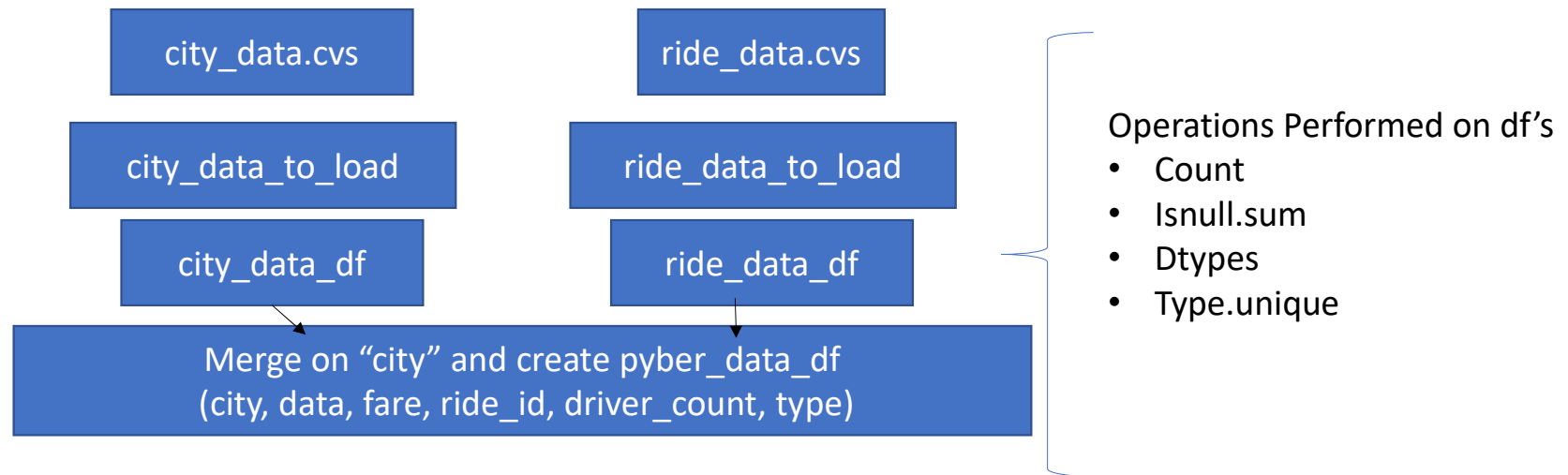
* No missing data was found in either df.

City Type DF's & Lists Created from pyber_data_df

Merge on "city" and create pyber_data_df
(city, data, fare, ride_id, driver_count, type)

Filter == Type
Urban_cities_df

.describe summary stats

Groupby city list(index)
urban_ride_count

.describe & mean, median, mode

Groupby city list(index)
urban_avg_fare

Groupby city list(index)
urban_driver_count

Filter == Type
suburban_cities_df

.describe summary stats

Groupby city list(index)
suburban_ride_count

.describe & mean, median, mode

Groupby city list(index)
suburban_avg_fare

Groupby city list(index)
suburban_driver_count

Filter ==Type
rural_cities_df

.describe summary stats

Groupby city list(index)
rural_ride_count

.describe & mean, median, mode

Groupby city list(index)
rural_avg_fare

Groupby city list(index)
rural_driver_count

Scatter Plot Combining All 9 Lists

(3 parameters: ride_count, avg_fares_driver_count)

Additional Lists and Stats Produced For Box and Whisker Plots

| Filter == Type Urban_cities_df	Filter == Type suburban_cities_df	Filter ==Type rural_cities_df
.describe summary stats	.describe summary stats	.describe summary stats
Create list urban_fares	Create list suburban_fares	Create list rural_fares
Calculate mean, median, mode	Calculate mean, median, mode	Calculate mean, median, mode
Create list urban_driver_count	Create list suburban_driver_count	Create list rural_driver_count
Calculate mean, median, mode	Calculate mean, median, mode	Calculate mean, median, mode
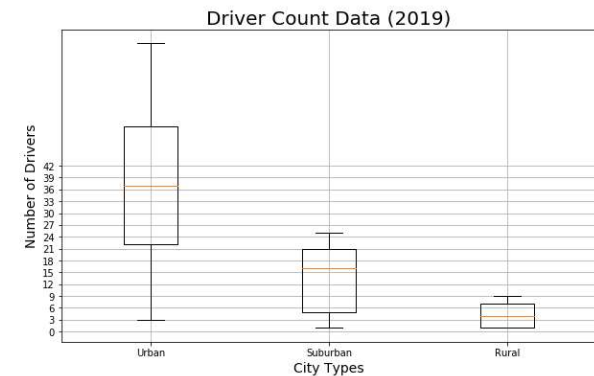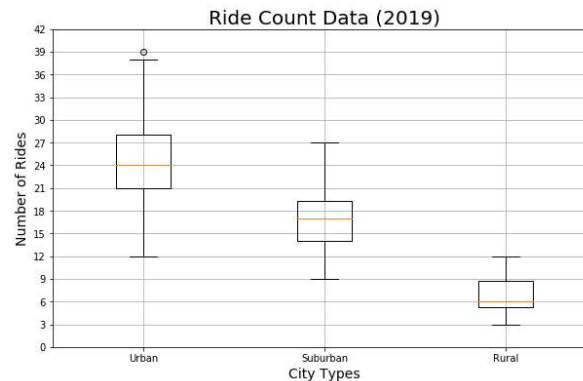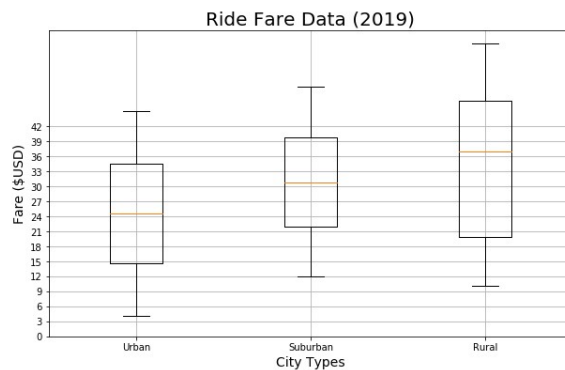
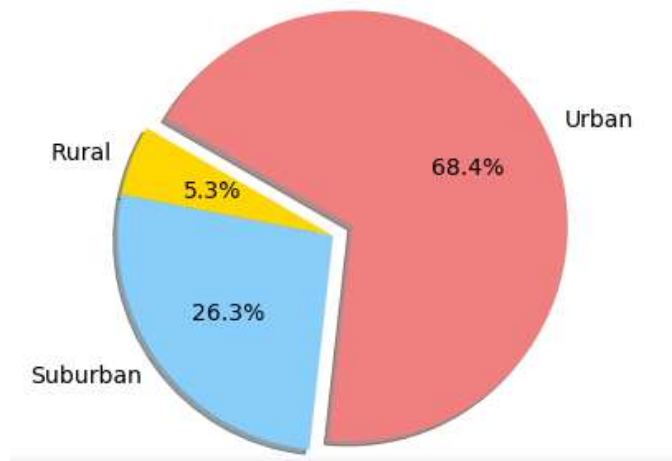# Groupby Lists and Calcs Produced Pie Charts
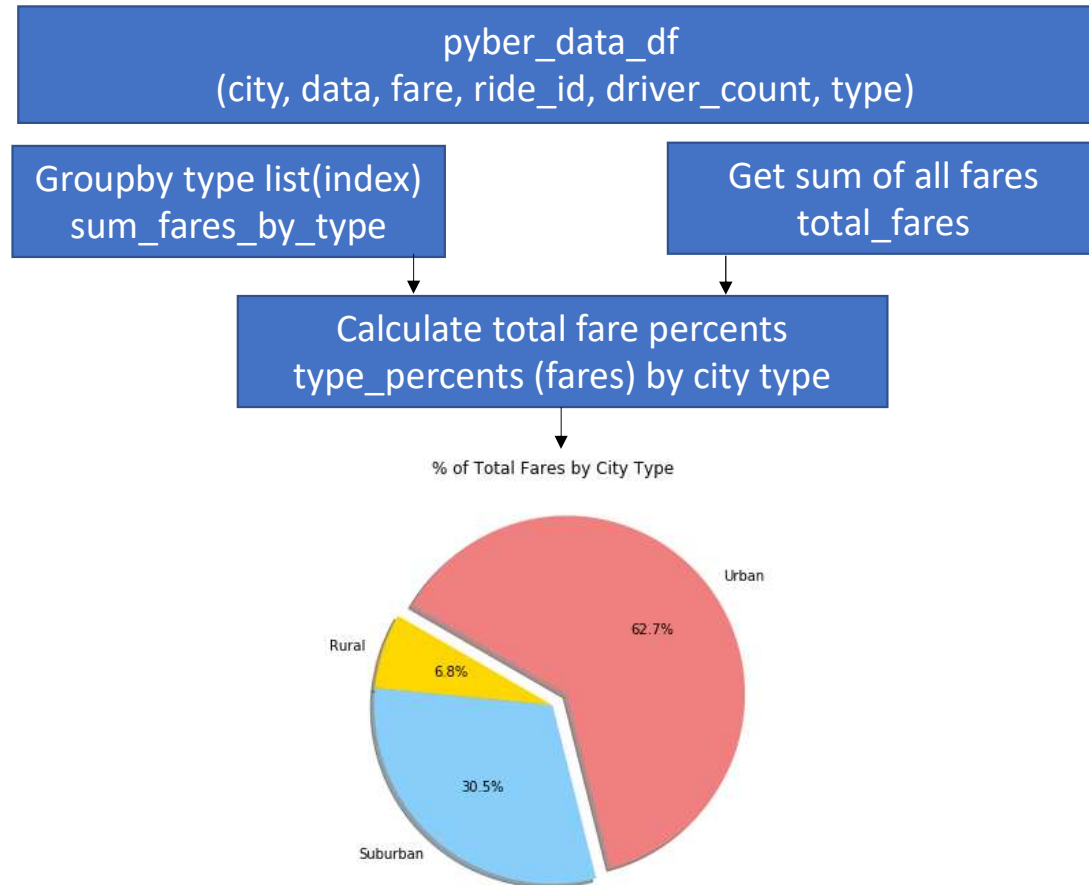
pyber_data_df
(city, data, fare, ride_id, driver_count, type)

Groupby type list(index)
ride_id_percents_by_type

Get sum of all ride_id
total_ride_id

Calculate percent by city type
ride_percents by city type
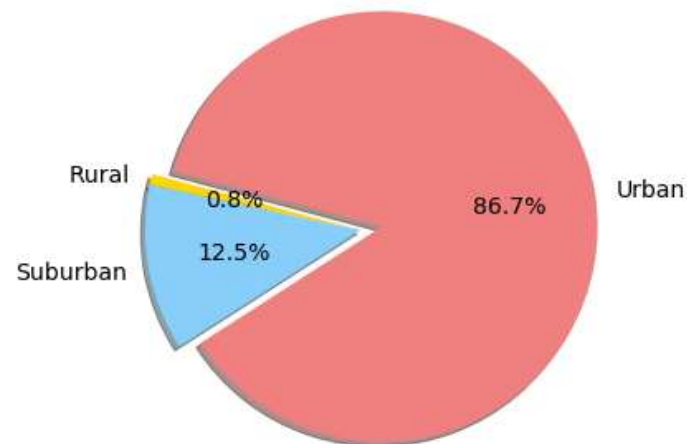
% of Total Rides by City Type

pyber_data_df
(city, data, fare, ride_id, driver_count, type)

Groupby type list(index)
driver_count by city type

Get sum of all drivers
driver_count

Calculate percent by city type
driver_percents by city type

% of Total Drivers by City Type

Challenge 5 – Summary DF