Karanjot Singh

In [130]:

```python
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
city_data = pd.read_csv("city_data.csv")
ride_data = pd.read_csv("ride_data.csv")
merge_data = ride_data.merge(city_data)
merge_data  = pd.DataFrame(merge_data)
#print(merge_data)

#Rural Data
rural = merge_data[merge_data['type']=='Rural']
#print(rural)
rural_mean = rural['fare'].groupby(rural['city']).mean()
#print(rural_mean)
rc = rural.groupby('city').count()
rural_count = rc.ride_id
#rural_count = rural['fare'].groupby(rural['city']).count()
#print(rural_count)
rural_driver_count = rural['driver_count'].groupby(rural['city']).count()
#plt.show()

#Urban Data
urban = merge_data[merge_data['type']=='Urban']
urban_mean = urban['fare'].groupby(urban['city']).mean()
uc = urban.groupby('city').count()
urban_count = uc.ride_id
#urban_count = urban['fare'].groupby(urban['city']).count()
urban_driver_count = urban['driver_count'].groupby(urban['city']).count()

#Suburban Data
suburban = merge_data[merge_data['type']=='Suburban']
suburban_mean = suburban['fare'].groupby(suburban['city']).mean()
sc = suburban.groupby('city').count()
suburban_count = sc.ride_id
#suburban_count = suburban['fare'].groupby(suburban['city']).count()
suburban_driver_count = suburban['driver_count'].groupby(suburban['city']).count()

#Plotting
plt.figure(figsize=(10,8))
plt.scatter(x = rural_mean,y = rural_count,s = rural_driver_count*50,color = 'b',alpha
plt.scatter(x = urban_mean,y = urban_count,s = urban_driver_count*50,color = 'r',alpha
plt.scatter(x = suburban_mean,y = suburban_count,s = suburban_driver_count*50,color =
plt.legend(['Rural','Urban','Suburban'],loc='best',title = 'City Type')
plt.xlabel('Average Fare')
plt.ylabel('Total Number of Rides')
plt.title('Number of Rides VS Total Number of Rides')
```
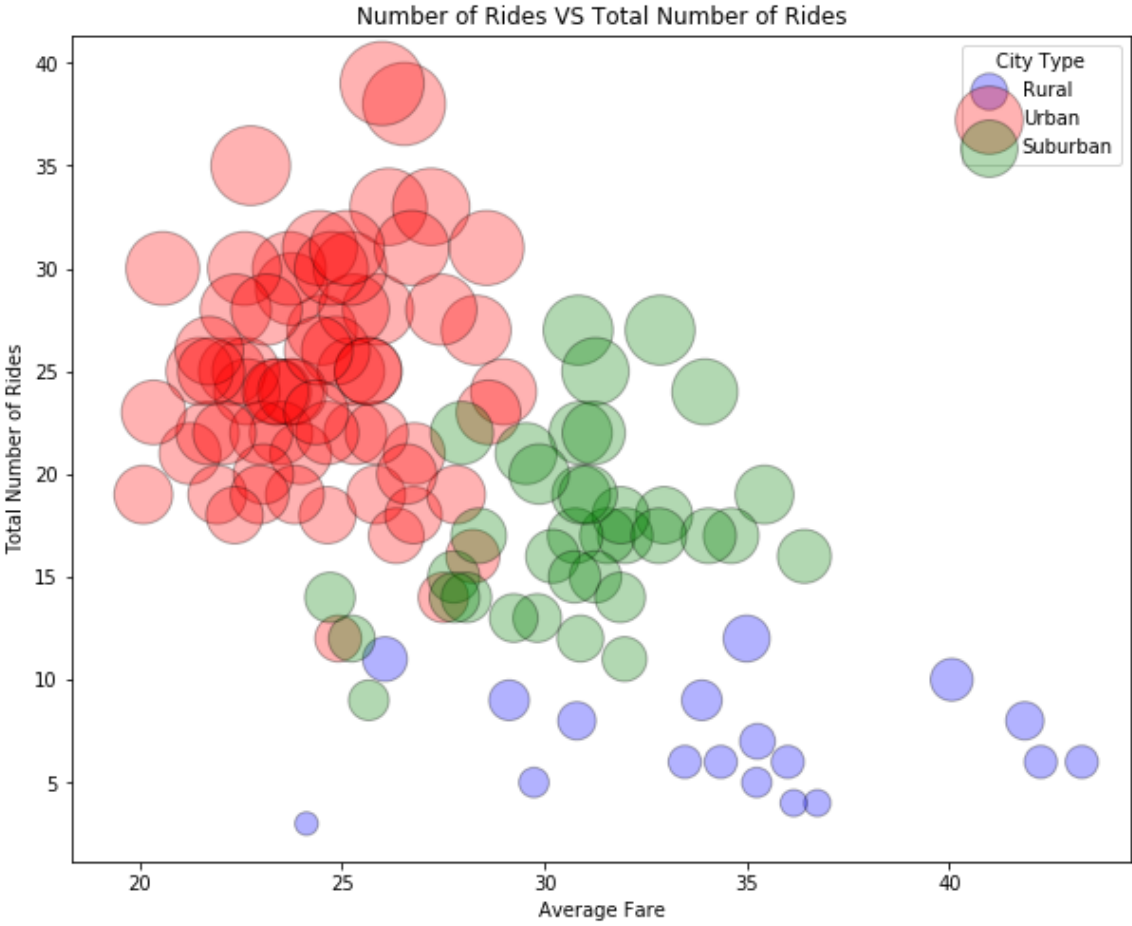
Out[130]:

Text(0.5, 1.0, 'Number of Rides VS Total Number of Rides')

Number of Rides VS Total Number of Rides

In [136]:

```python
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
rural_sum = rural['fare'].sum()
urban_sum = urban['fare'].sum()
suburban_sum = suburban['fare'].sum()
lab = ['Rural','Urban','Suburban']
Sum = [rural_sum,urban_sum,suburban_sum]

fid = plt.figure(figsize=(10,5))
ax1 = plt.subplot()
ax1.pie(Sum,labels = lab,autopct = '%1.1f%%',shadow = True,startangle =90)
plt.title('Total Fare By City Type')
plt.show()
```
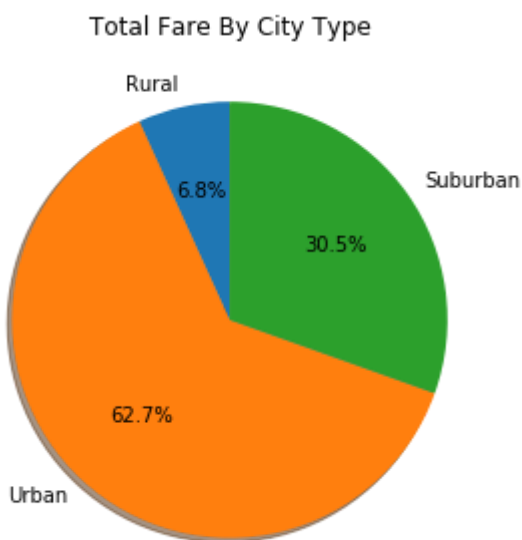
In [135]:

```python
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
total_rural_rides = rural['ride_id'].count()
total_urban_rides = urban['ride_id'].count()
total_suburban_rides = suburban['ride_id'].count()
total_rides = [total_rural_rides,total_urban_rides,total_suburban_rides]
plt.figure(figsize=(10,5))
ax1 = plt.subplot()
ax1.pie(total_rides,labels = lab,autopct = '%1.1f%%',shadow = True,startangle =90)
plt.title("Total Rides by City Type")
plt.show()
```
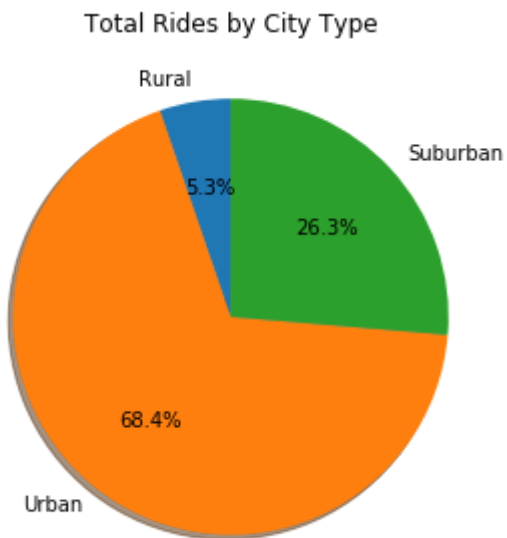
In [139]:

```python
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
total_rural_rides = rural['driver_count'].sum()
total_urban_rides = urban['driver_count'].sum()
total_suburban_rides = suburban['driver_count'].sum()
total_rides = [total_rural_rides,total_urban_rides,total_suburban_rides]
plt.figure(figsize=(10,5))
ax1 = plt.subplot()
ax1.pie(total_rides,labels = lab,autopct = '%1.1f%%',shadow = True,startangle =90)
plt.title("Total Driver By City Type")
plt.show()
```

Total Driver By City Type