DATA STRUCTURES

Group Members:
M.Shahzaib (20k-1067)
Sohaib Ashraf (20k-0488)
Khawaja Abdullah (20k-0385)

Miss Mubashra

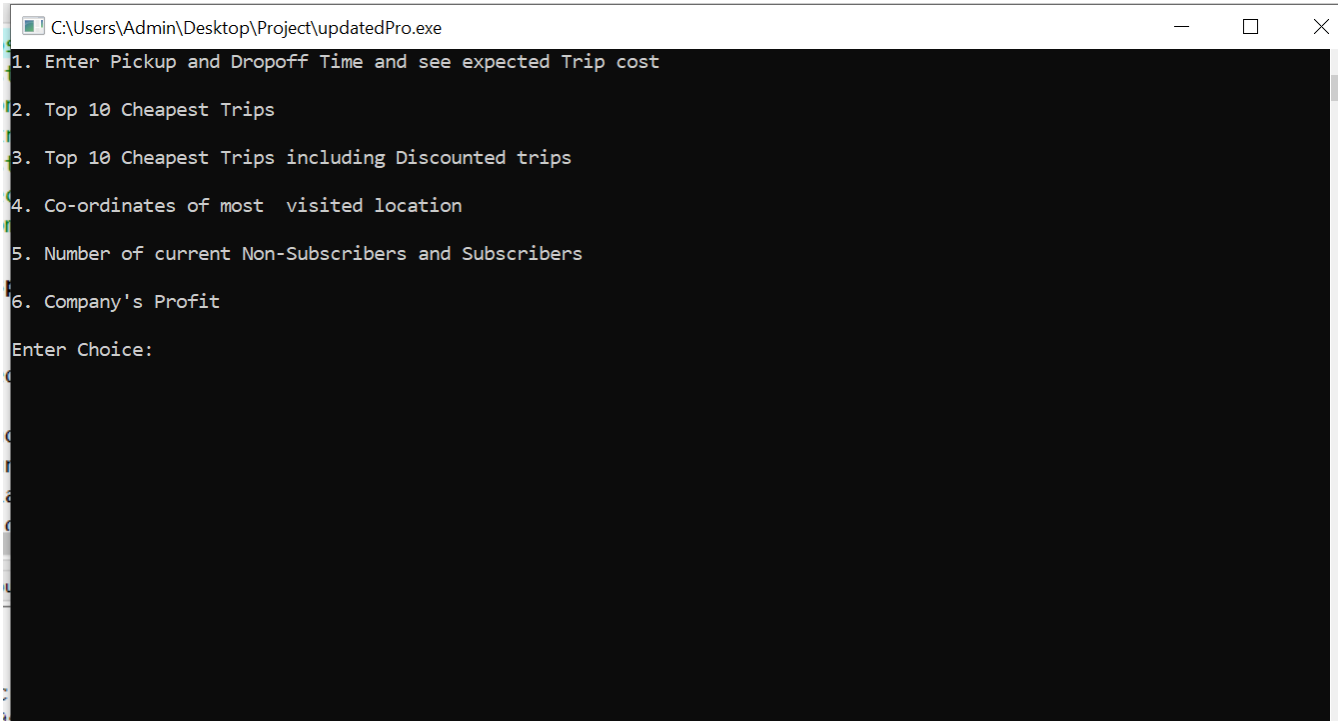
The Taxi Trip[Report]

Dataset:

The data set used in this project is of Yellow taxi trip data by New York City (NYC) Taxi & Limousine Commission (TLC). These are the famous NYC yellow taxis that provide transportation exclusively through street hails.
Source: https://www.kaggle.com/elemento/nyc-yellow-taxi-trip-data?select=yellow_tripdata_2016-01.csv
Dataset size: 1.71 GB

AIM:

The main objective was to use the data available in our datasets to analyze the number of rides and calculate things like peak factor, number of subscribers and non-subscribers, top cheapest rides, top discounted rides, most visited location and total profit of both company and the drivers.



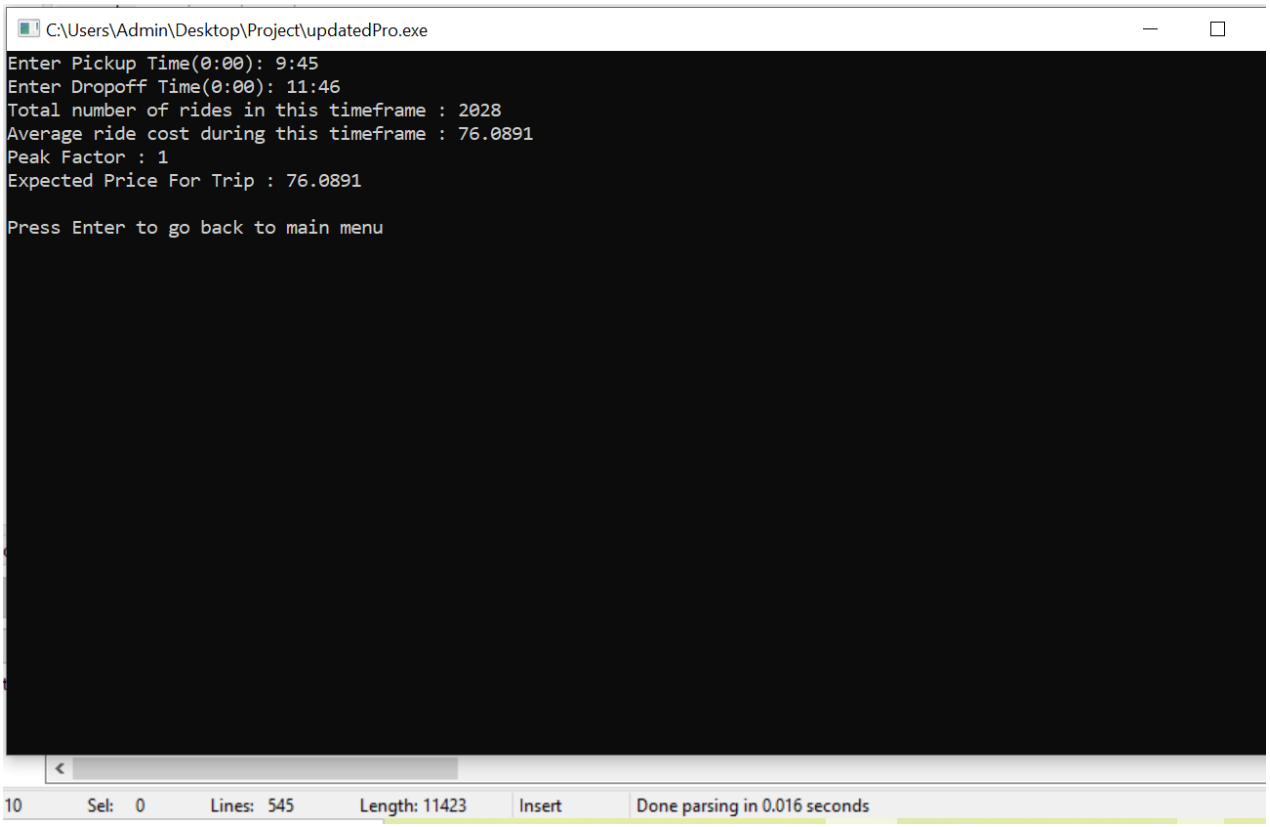
Procedure:

We first read the data of our CSV file and store them in variables and construct a Binary Search Tree based on the total fare amount.

We filtered the data and removed the rides with more than 4 passengers. We also removed some outliers from the coordinates and only kept coordinates within the New York City boundary.

Calculating Peak Factor:

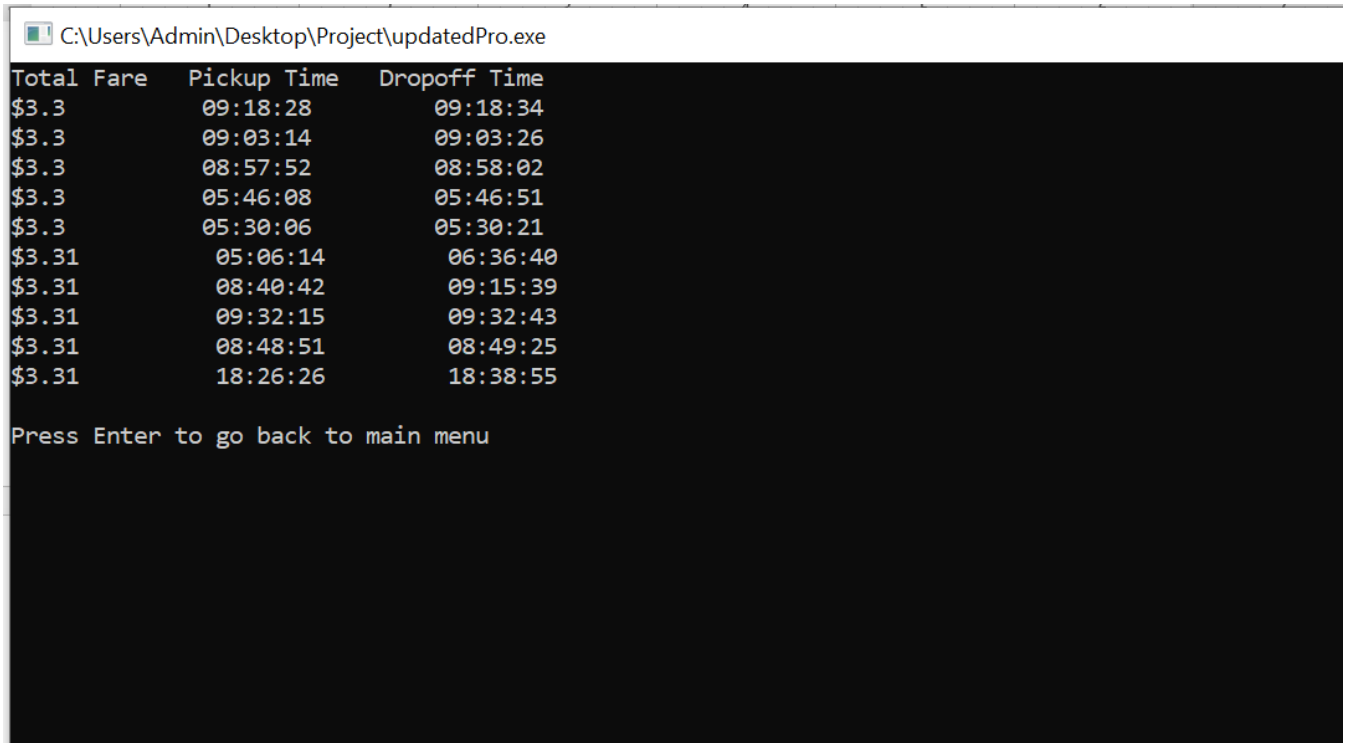
- We take pickup and drop off time from the user as input. The trip cannot be of more than 12 hours.
- We also validate user's input so we get the data in the correct format otherwise the user is asked to enter the input again.
- We check the rides between this time slot in our binary tree and calculated the peak factor based on it.
- The output shows the peak factor and estimated fare amount



```
tree.inorder(tree.root,phour,pmin,dhour,dmin);
cout<<"Total number of rides in this timeframe : "<<tree.count<<endl;
cout<<"Average ride cost during this timeframe : "<<tree.sum/tree.count<<endl;
int pf = (tree.sum/tree.count);
pf = pf%3;
if(pf == 0){
pf = 1;
}
cout<<"Peak Factor : "<<pf<<endl;
cout<<"Expected Price For Trip : "<<(tree.sum/tree.count)*pf<<endl;
cout<<endl<<"Press Enter to go back to main menu"<<endl;
getch();
break;
```

Top cheapest rides:

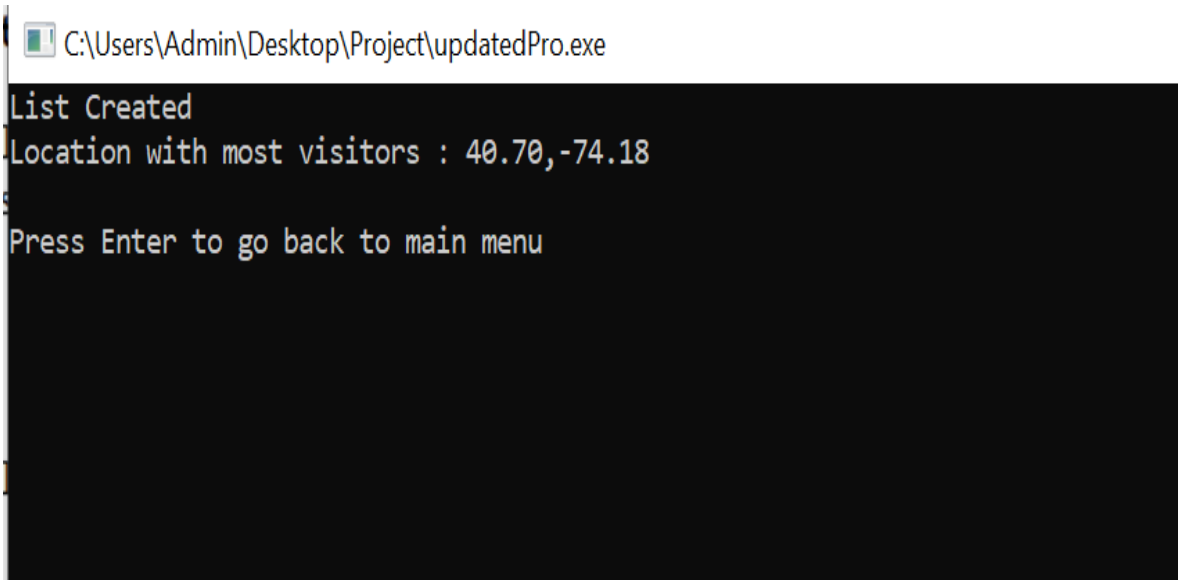
We used the fare amount and compare them and took out top ten cheapest rides with their pickup and drop off time. The negative amount also indicates that the customer had credit in his account or he had discount vouchers which makes the ride cheaper than others. We calculated both separately(the ones with discount and ones without discount).



Coordinates of most visited location:

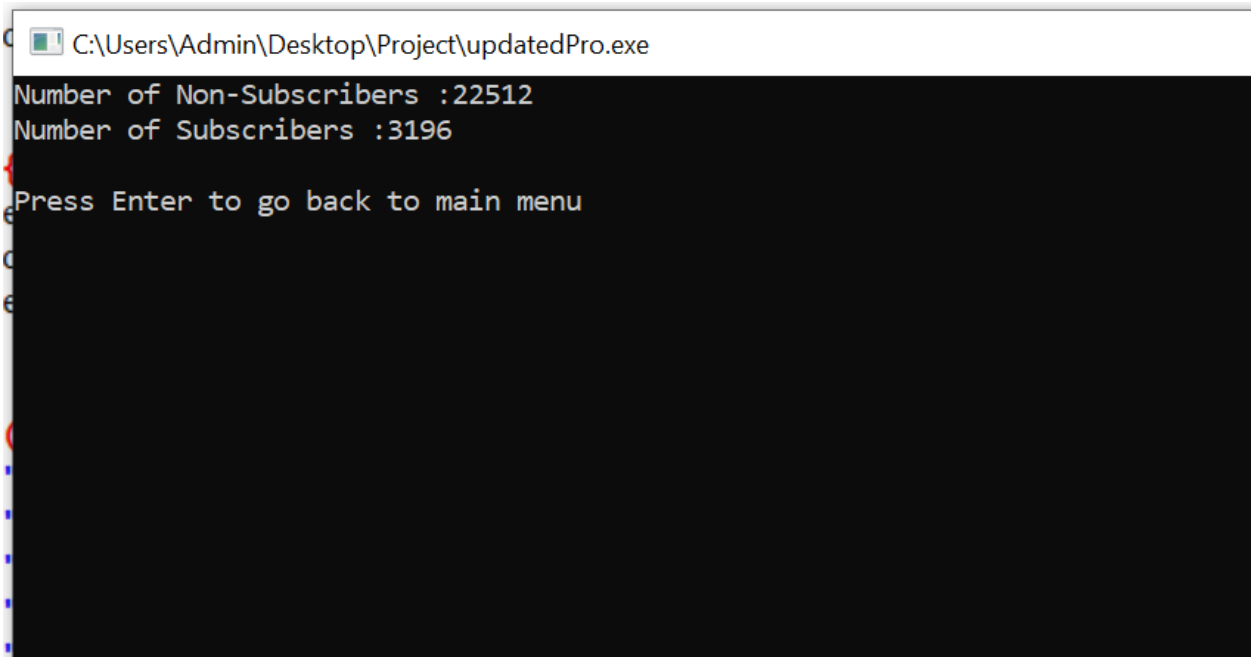
Next we found out coordinates of most visited place of New York using our dataset. We created singly linked list of all drop off coordinates. If the coordinates are matched, we increased the count of that cord. Next we found out the max count and output the coordinates with maximum count which tells us the most visited location of New York.

```
bool search(double lat,double lon){
    if(head==NULL){
        return false;
    }
    linkednode *temp = head;
    while(temp!=NULL){
        if(temp->latitude==lat && temp->longitude==lon){
            temp->count++;
            return true;
        }
        temp  = temp->next;
    }
    return false;
}
void checkMax(){
    linkednode *temp = head;
    int max=0;
    while(temp!=NULL){
        if(max<temp->count){
            max= temp->count;
        }
        temp= temp->next;
    }
```



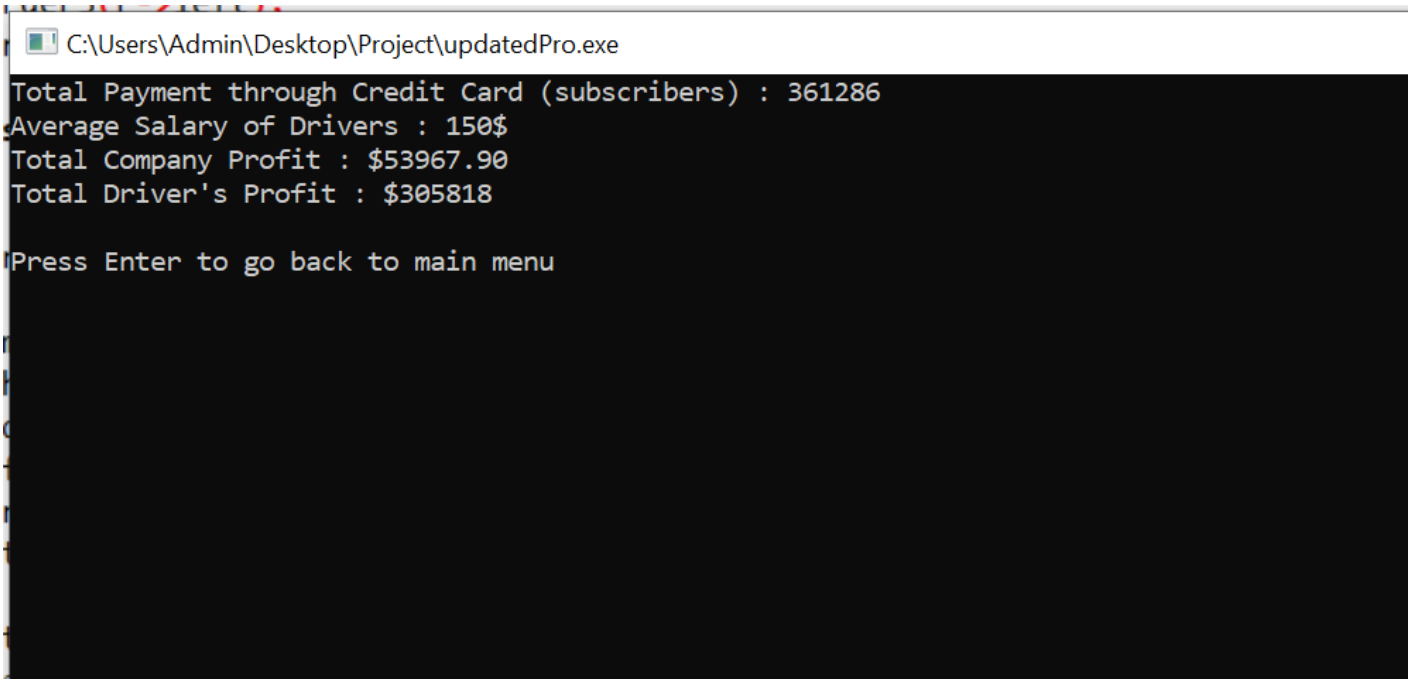
Number of Subscribers and Non-Subscribers:

Our dataset contains a column of payment type. We used it effectively to calculate out total number of subscribers (who paid from our card) and non-subscribers who paid with cash. We will also use this information later to calculate our company and drivers' profit.



Company And Drivers' Profit:

- We filtered all the rides where payment was made by card (our subscribers)
- We added the total amount to calculate company's profit.
- We kept drivers salary as $150. The profit is calculated and 15% is deducted from drivers' profit and added to company's profit as company's commission and profit.


C:\Users\Admin\Desktop\Project\updatedPro.exe
Total Payment through Credit Card (subscribers) : 361286
Average Salary of Drivers : 150$
Total Company Profit : $53967.90
Total Driver's Profit : $305818
Press Enter to go back to main menu