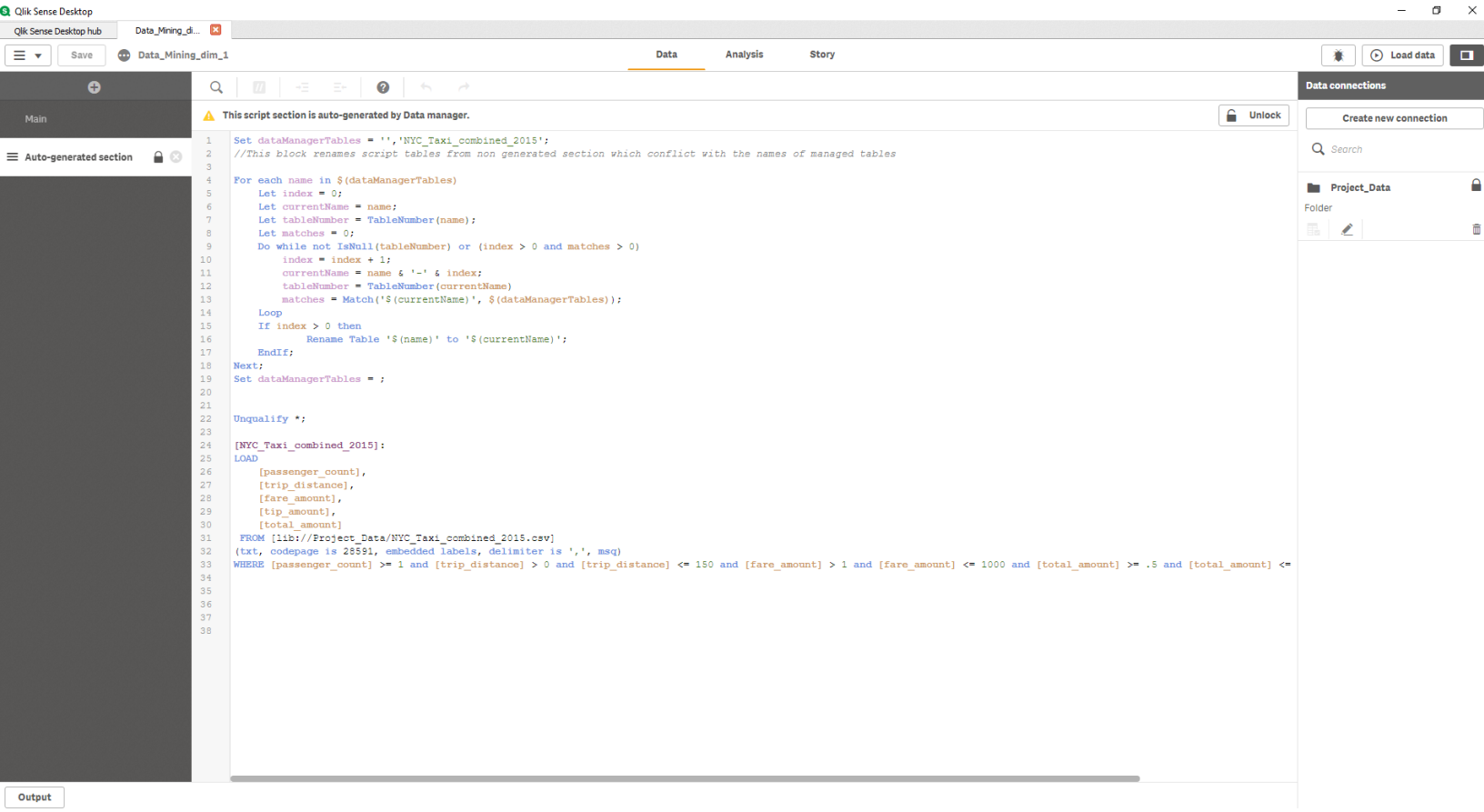
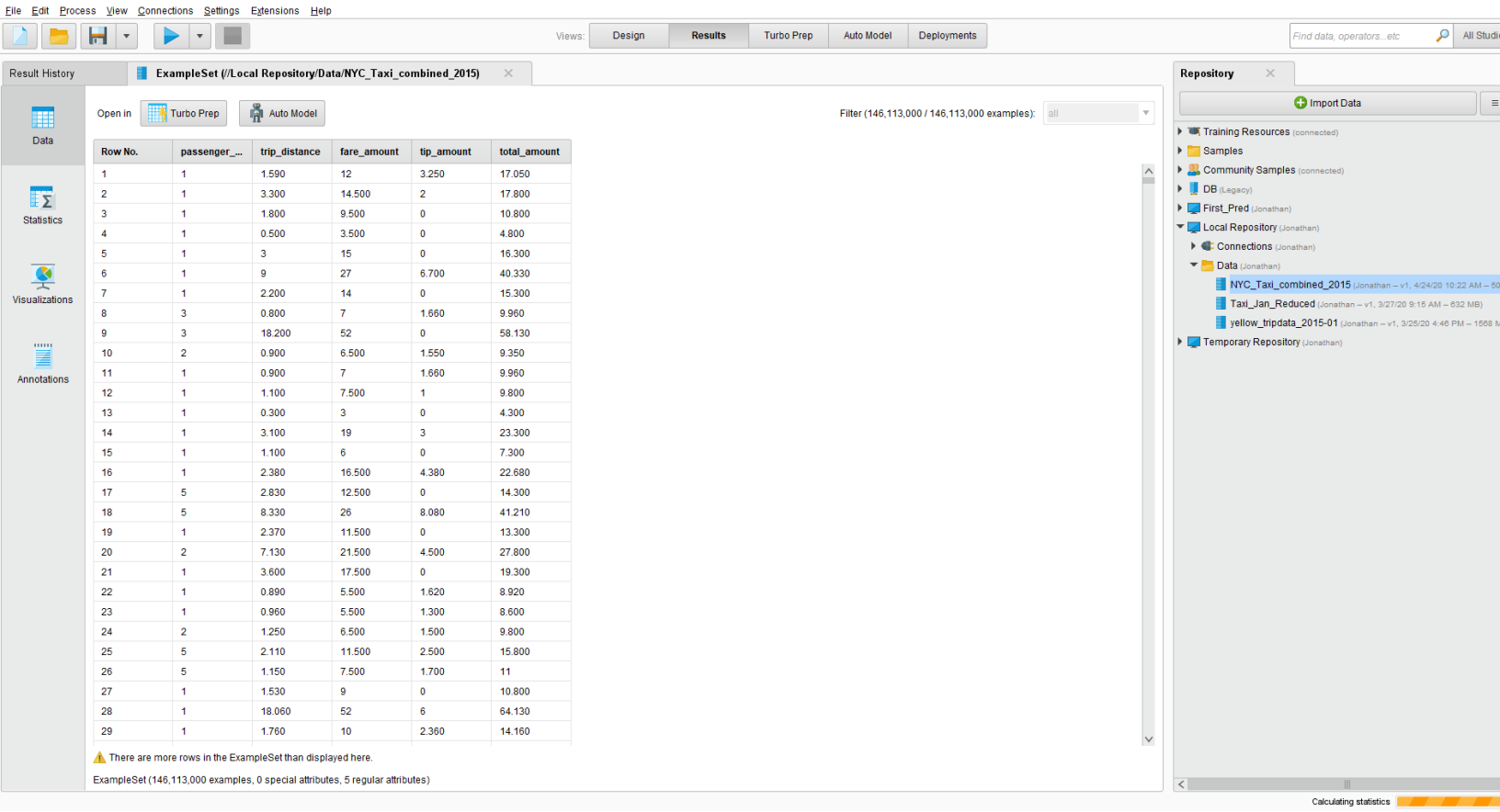
Overall, I used various tools for my analysis as opposed to crafting original code. Below is the script I used for Qlik, which was auto-generated. This section essentially shows the data being loaded into Qlik.




```
Set dataManagerTables = '','NYC_Taxi_combined_2015';
//This block renames script tables from non generated section which conflict with the names of managed tables

For each name in $(dataManagerTables)
    Let index = 0;
    Let currentName = name;
    Let tableNumber = TableNumber(name);
    Let matches = 0;
    Do while not IsNull(tableNumber) or (index > 0 and matches > 0)
        index = index + 1;
        currentName = name & '-' & index;
        tableNumber = TableNumber(currentName)
        matches = Match('$(currentName)', $(dataManagerTables));
    Loop
    If index > 0 then
            Rename Table '$(name)' to '$(currentName)';
    EndIf;
Next;
Set dataManagerTables = ;


Unqualify *;

[NYC_Taxi_combined_2015]:
LOAD
    [passenger_count],
    [trip_distance],
    [fare_amount],
    [tip_amount],
    [total_amount]
 FROM [lib://Project_Data/NYC_Taxi_combined_2015.csv]
(txt, codepage is 28591, embedded labels, delimiter is ',', msq)
WHERE [passenger_count] >= 1 and [trip_distance] > 0 and [trip_distance] <= 150 and [fare_amount] > 1 and [fare_amount] <= 1000 and [total_amount] >= .5 and [total_amount] <=
```

Included in this section is also a javascript file called "codewander-k-means-cluster." This was the code used for the k-means clustering via a Qlik extension. Neither the code, nor the entirety of the extension, are my own. I utilized them as a tool to explore the data.

Another tool that was utilized was RapidMiner. Some of the data was uploaded and transformed via RapidMiner



Unfortunately, RapidMiner restricted access to their code so I am unable to provide any descriptions of it. However, given how useful this tool was I wanted to include a screencap of it. One of the problems we faced with this project was being able to provide enough RAM and other resources to properly analyze the data with certain tools. With snippets of the data I was able to transform, clean, and analyze certain pieces and then analyze in large scale in different ways. If I had better resources I think this would have been my go to tool.