Ty’s Data Work Flow Example

Author: Ty Tanita

Date: 4/17/19

Note:

This can all be done with CSVs too, so if you have big data, it is fine.

Future thoughts:

Create an interphase to enter data and have it populate in the backend. This could be a GUI in Excel or use JavaScript and make a site. Then I could also create survey questions, and have it fill out the backend sheet for loading.

This could also be hosted in AWS. For this example, I ran everything off my laptop.

While all these tools already exist, I try to explain these concepts in a relevant and easy to understand way so others can find ways to apply these concepts as well. In this example, I show each tool and how they are interconnected.

Table of Contents

[Raw Data 3](#_Toc1141862)

[MySQL 4](#_Toc1141863)

[Python 10](#_Toc1141864)

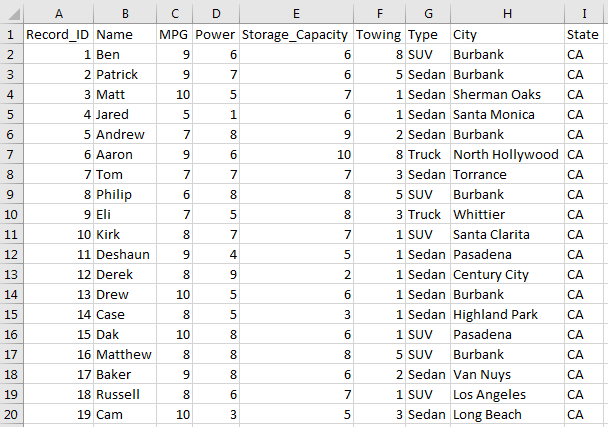
[Tableau 19](#_Toc1141865)

[AI/ML 26](#_Toc1141866)

[Ty’s Data Tool Chart 34](#_Toc1141867)

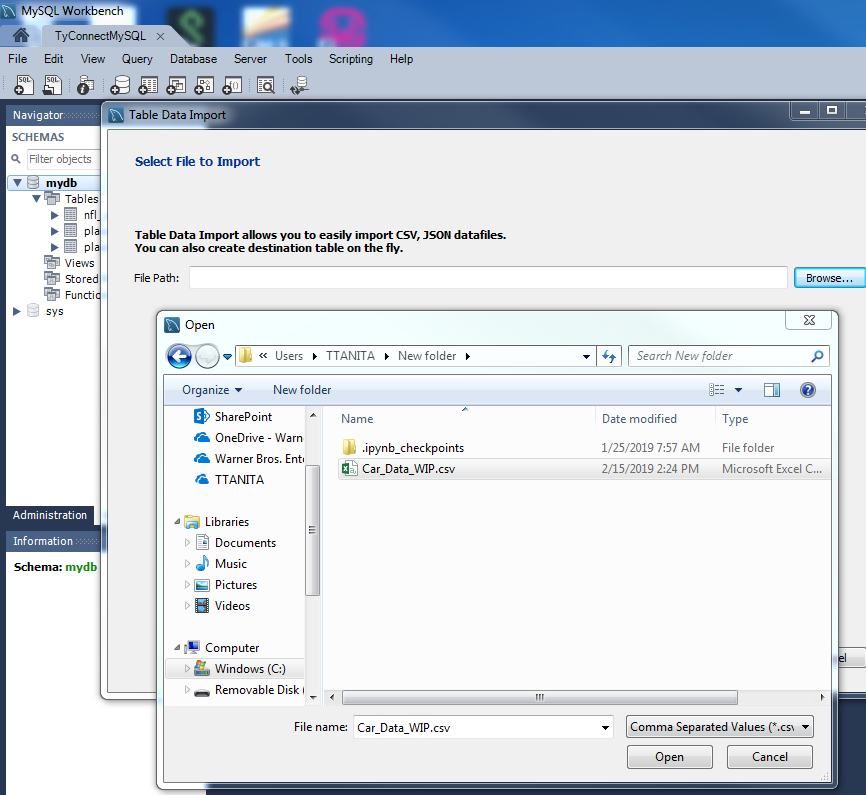
# Raw Data

Here is our raw data:

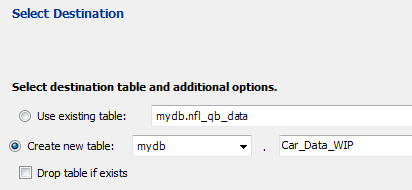


# MySQL

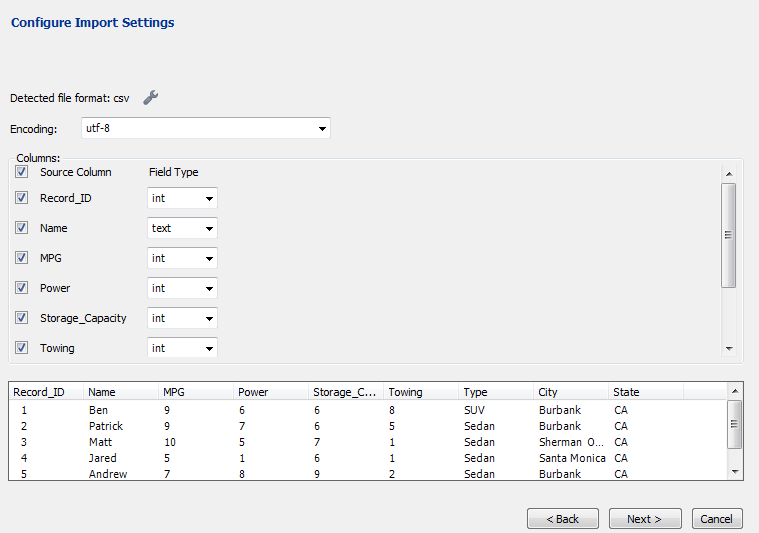
Now let’s load it into MySQL. This should be a CSV file which is fine for big data:



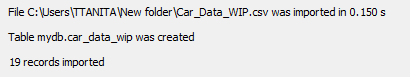
Let’s call this the, “Car\_Data\_WIP” table:



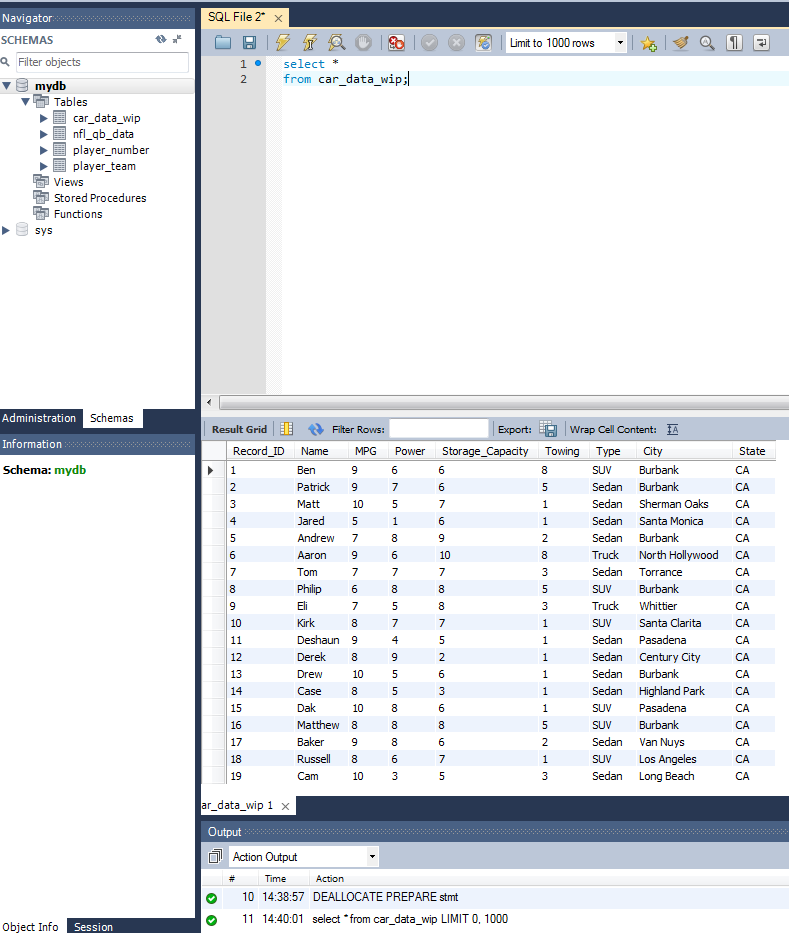
Setup the data types:



The records were successfully imported, if you are doing more data, you can get an idea of how long it would take:



Let’s run a query to see our data:

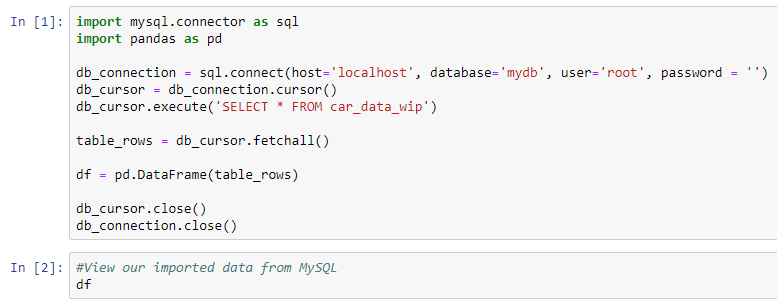


# Python

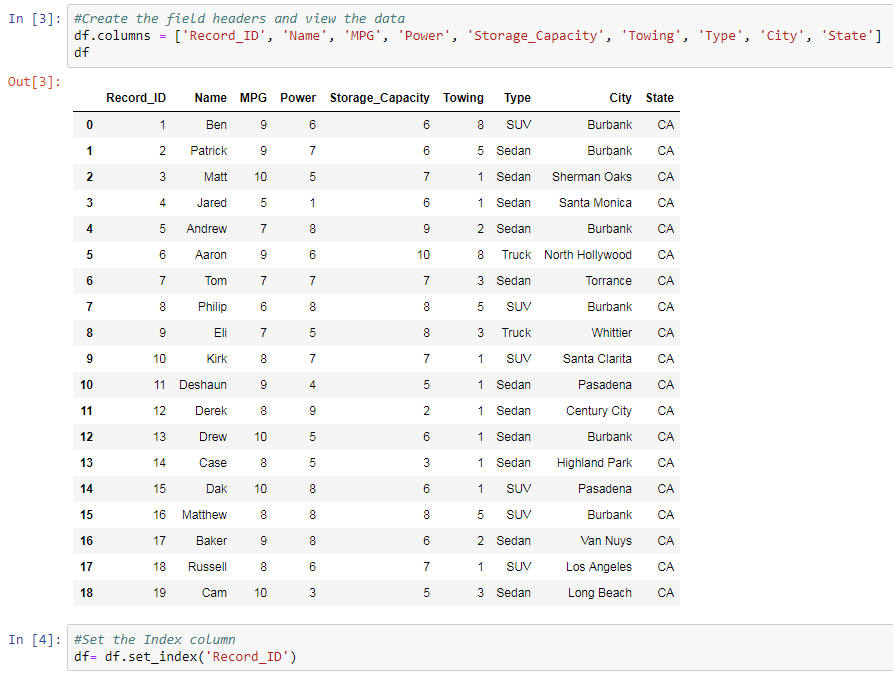
Now that our data is in our data base, let’s use Python to import and edit our data. This would also be helpful step if we need to clean big data. We could import it, run some scripts and then export it back to the database.

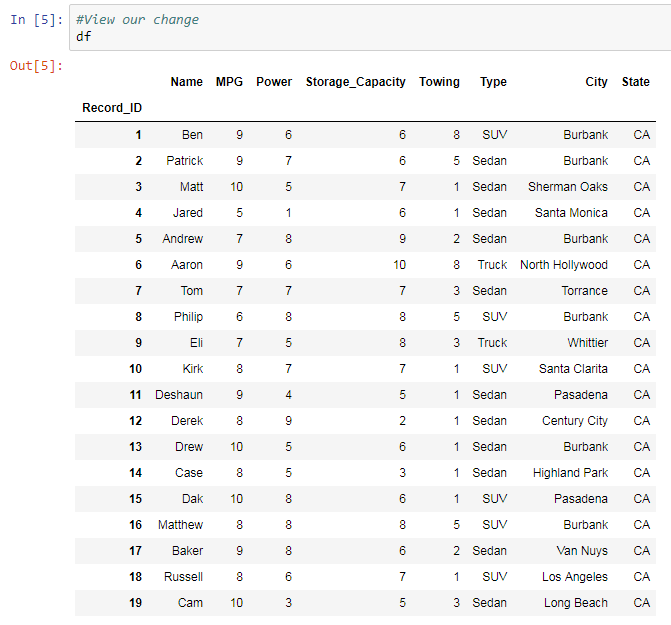
I have this as a separate file, so I will only show the screen shot in this document:

<https://github.com/tygithub18/Python_and_MySQL/blob/master/How%20to%20import%20and%20edit%20data%20from%20a%20database%20with%20Python.ipynb>

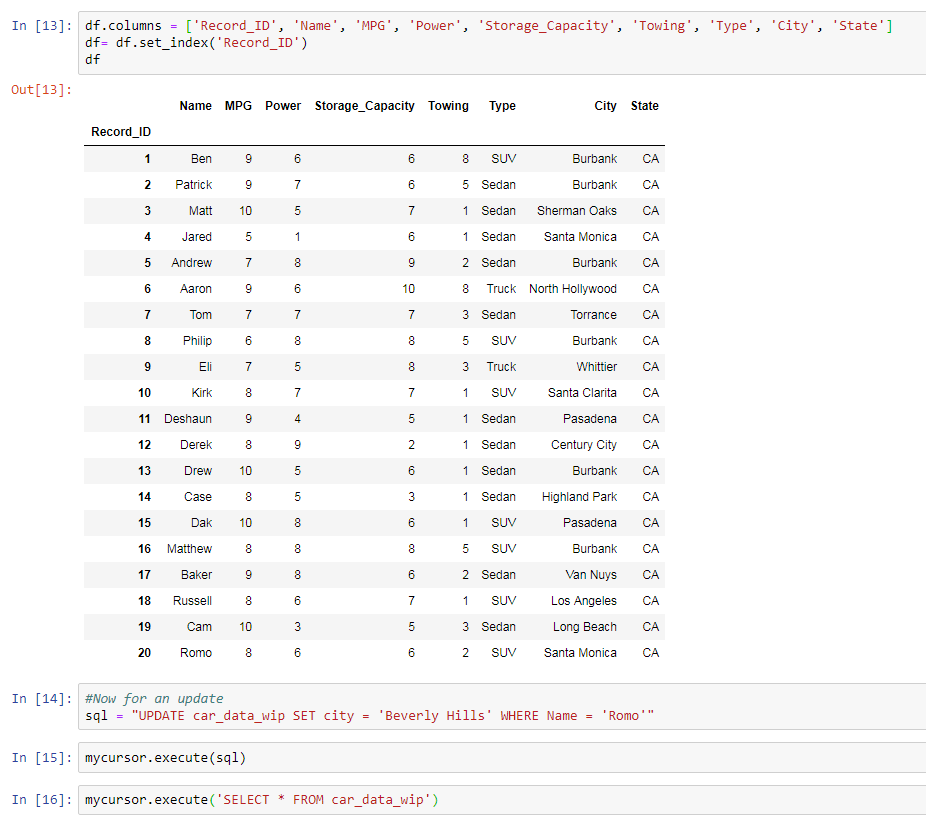


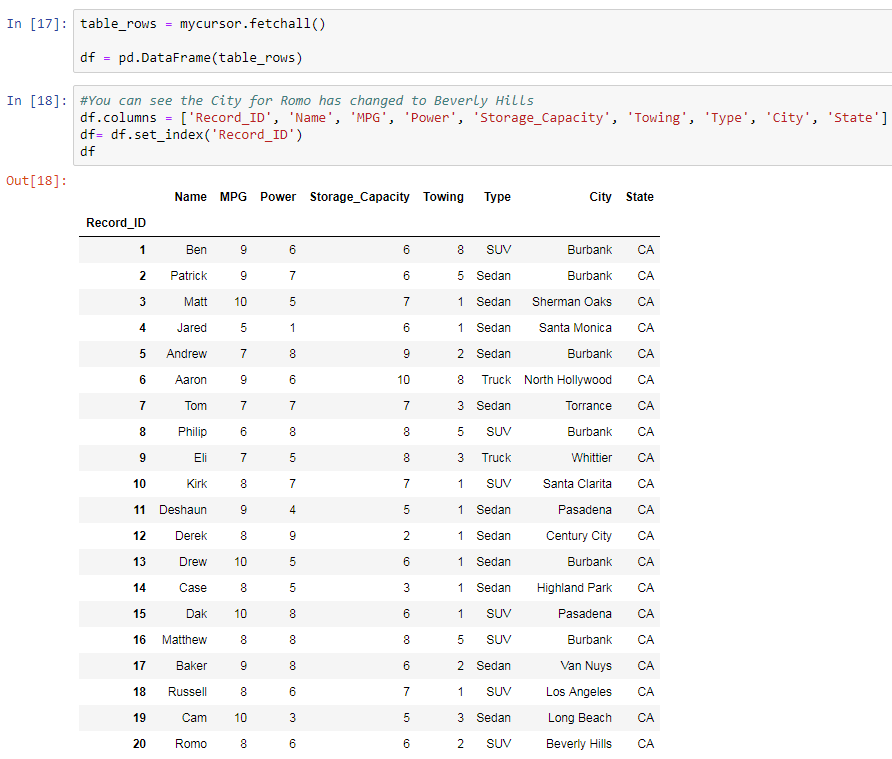


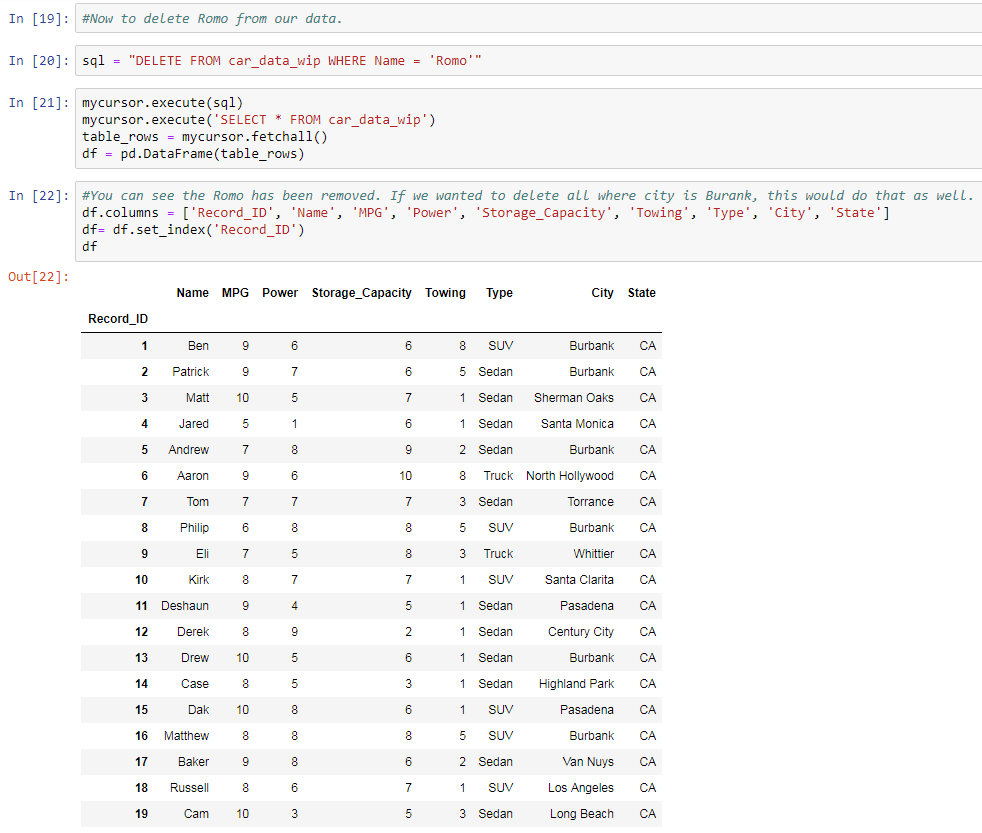


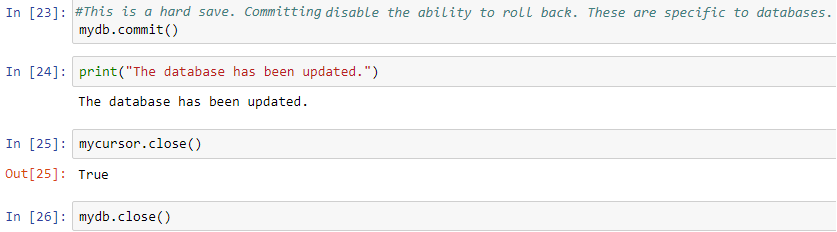










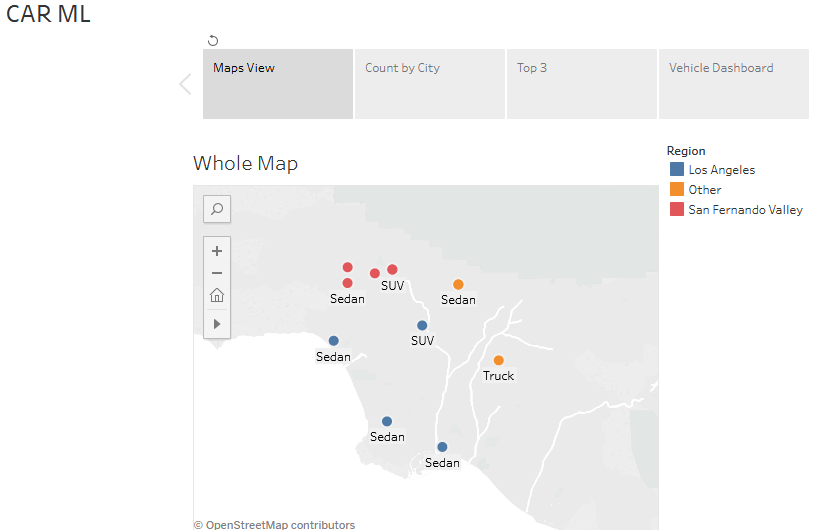


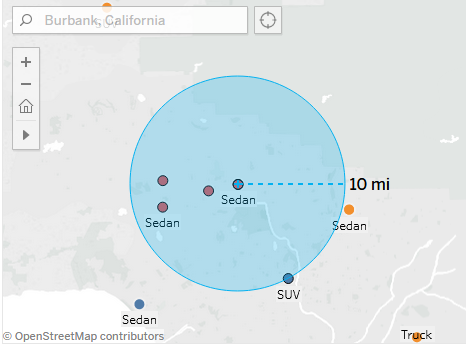
# Tableau

Now that our data is setup in MySQL, lets create some data visuals with Tableau

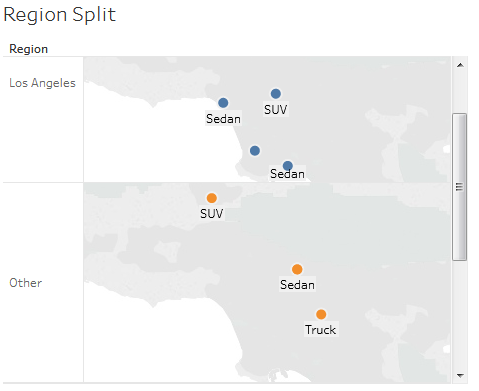
<https://public.tableau.com/profile/ty6116#!/vizhome/MLCarSample/CarML?publish=yes>

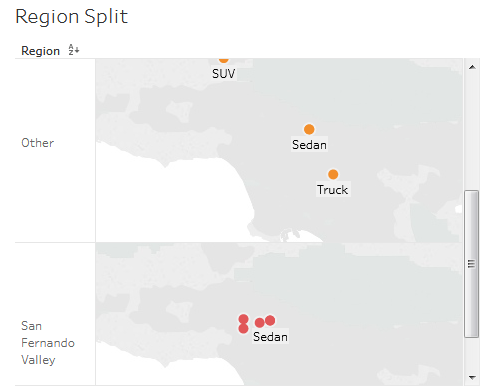
Here is a map view where we can see how many people are within 10 miles of Burbank.



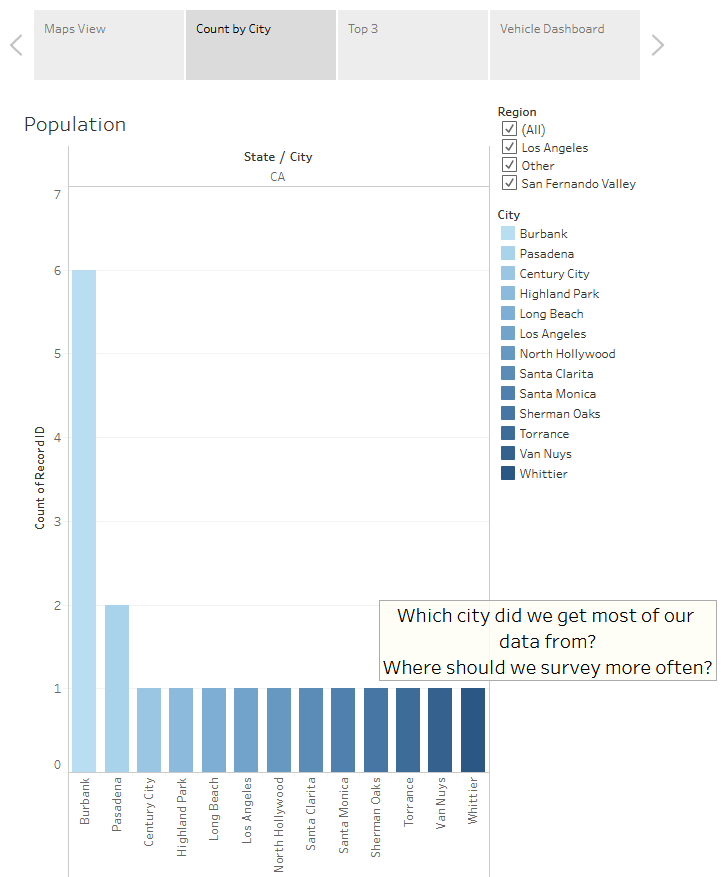


This is a custom region split view I created to help identity where people are from:

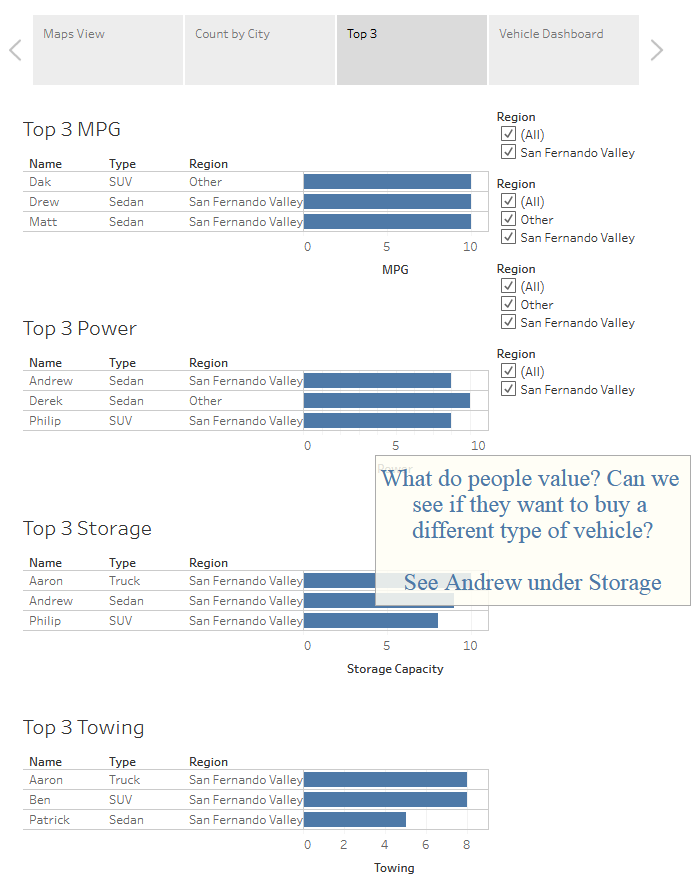




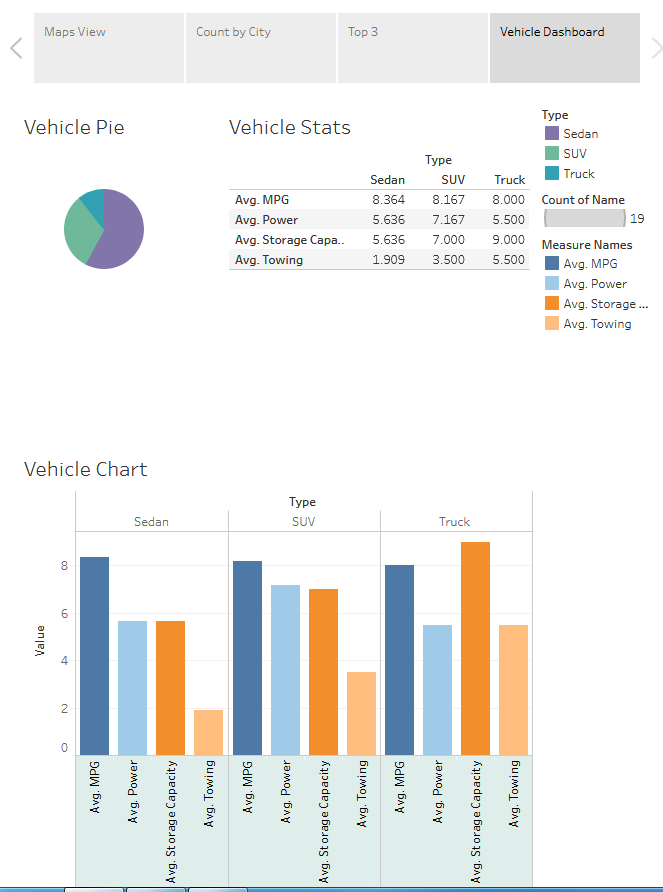
This shows a count of the number of people from each city:



Here is a view of the top three per vehicle attribute. Notice that we might be able to see Andrew a SUV or Truck.



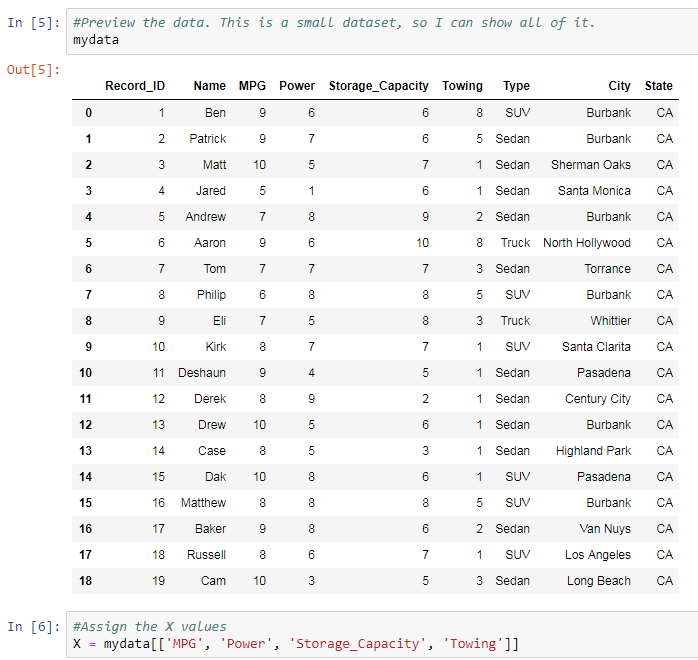
You can have multiple charts of related and non-related information at the same time:

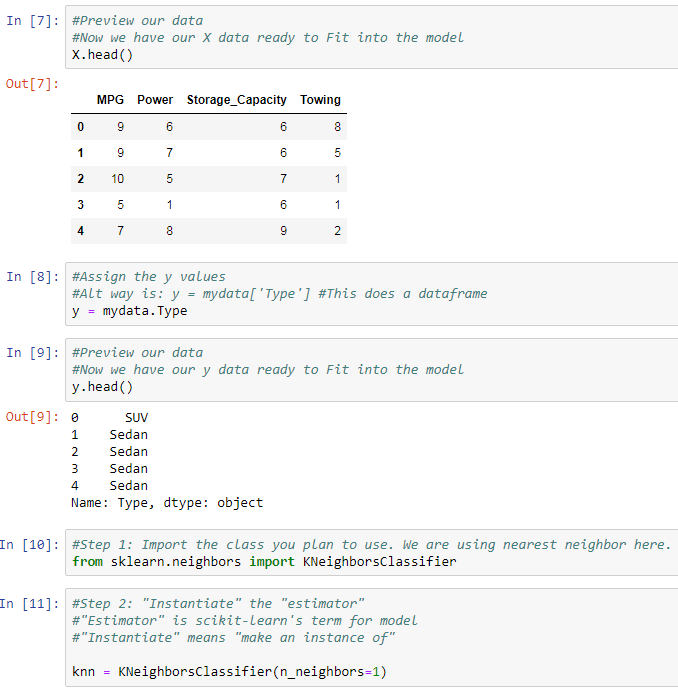


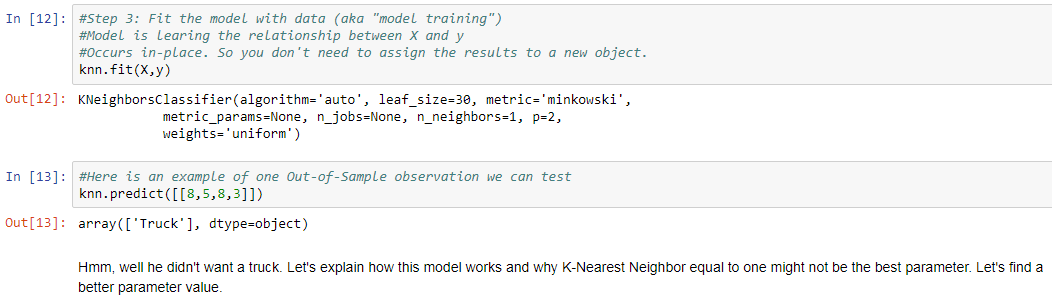
# AI/ML

This is my model that I created using Scikit-learn:

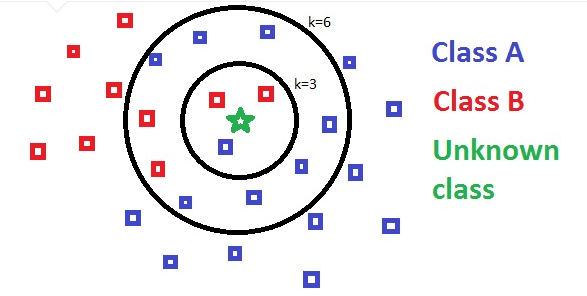




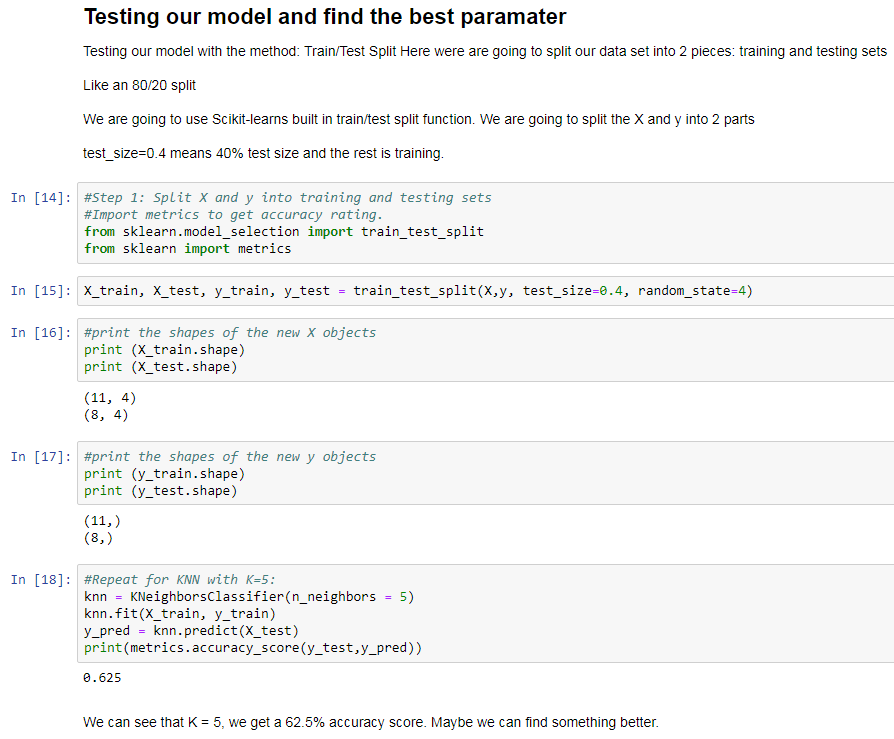


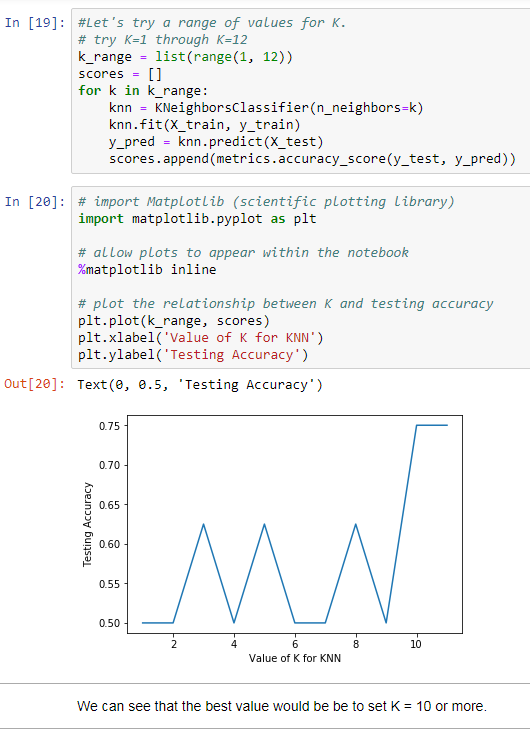


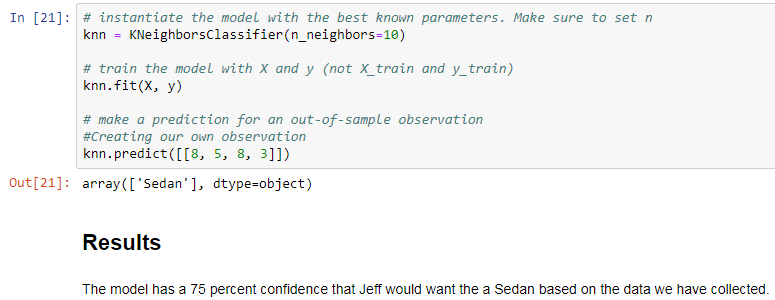
This is how K Nearest Neighbors works:



It looks at what values are closest to your unknown value on a scatter plot. Then it takes the majority of the values to make its prediction.







# Ty’s Data Tool Chart

This is a data flow chart that I created to help identify what stage in the data process a current project is at. This also helps visualize to stakeholders what their situation is and allow them to see the steps going forward. In addition, I can communicate with the technical regarding various tools what present the work flow model. From a student’s perspective, this creates a clear path because there are too many tools that appear in a Google search. While the tools may change, the process will always exist.

