Gathering Formula1 (F1) Data for this project:

My initial goal for this project was to see what effect major rule changes had or (did not have) on F1 viewership. After spending a few days trying to get viewership data, I decided that that may not be possible. I could manually find data for each year (and I started to do that) but with the changes to F1 ownership and television rights changes. I did not believe that the numbers would prove much, since there are significant changes on broadcasters from year to year and recently a pay to play option was introduced. Meaning that there may be less viewers but F1 owners may make more money in the long run. After doing this research I decided to change course and instead focus on F1 rule changes and its effect on race competitiveness.

When searching for F1 historical data, I found the Ergast Developer API website (https://ergast.com/). This website provides F1 historical data from 1950 to today. It has data on races, drivers, results, etc. The website offers several ways to access the data, via csv files, via downloading the schema, and via an RESTful API. I downloaded the schema to MySql and began looking into creating my data set from the F1 tables.

Even though I'm a lot more comfortable working with SQL queries to retrieve data, I decided to gather the data using the API. The API returns JSON data . Depending on the parameters you pass to the URL the GET requests returns data in different formats. You can filter the data down to a driver and specific race. For this project I needed historical data for all races. I also decided that the top three spots for each race.

To retrieve this data I used the url that like this: "https://ergast.com/api/f1/2012/results/1.json". Where the '2012' is the year of the race and the '1' at the end is the race result. This returns nested JSON output (see sample below). The results were relatively easy to access, but the driver and constructor details were in a nested dictionary.

Once I was able to traverse through the JSON output to retrieve the data I wanted, I created a function to loop through the data and export it to a CSV file. The function can be modified to retrieve data for different date ranges (years) and different results. I will be using data from 1980 to 2018 and for the top 3 finishes in each race.

Once the data is exported to a CSV file it can easily be turned into a Pandas dataframe (see sample below). The historical data contains the following fields:Posittion, RaceName, SeasonYear, RaceNumber, RaceDate, DriverFirstName, DriverLastName, DriverID, StartingGridNumb, ConstructorName.

I will use information in the following Wikipedia page to compost the list of years/dates that had significant rule changes:http://en.wikipedia.org/wiki/History_of_Formula_One_regulations.

Sample JSON Result:

{'season': '2012',
 'position': '1',
 'Races': [{'season': '2012',
   'round': '1',
   'url': 'http://en.wikipedia.org/wiki/2012_Australian_Grand_Prix',
   'raceName': 'Australian Grand Prix',
   'Circuit': {'circuitId': 'albert_park',
    'url': 'http://en.wikipedia.org/wiki/Melbourne_Grand_Prix_Circuit',
    'circuitName': 'Albert Park Grand Prix Circuit',
    'Location': {'lat': '-37.8497',
     'long': '144.968',
     'locality': 'Melbourne',
     'country': 'Australia'}},
   'date': '2012-03-18',
   'time': '06:00:00Z',
   'Results': [{'number': '3',
     'position': '1',
     'positionText': '1',
     'points': '25',
     'Driver': {'driverId': 'button',
      'permanentNumber': '22',
      'code': 'BUT',
      'url': 'http://en.wikipedia.org/wiki/Jenson_Button',
      'givenName': 'Jenson',
      'familyName': 'Button',
      'dateOfBirth': '1980-01-19',
      'nationality': 'British'},
     'Constructor': {'constructorId': 'mclaren',
      'url': 'http://en.wikipedia.org/wiki/McLaren',
      'name': 'McLaren',
      'nationality': 'British'},
     'grid': '2',
     'laps': '58',
     'status': 'Finished',
     'Time': {'millis': '5649565', 'time': '1:34:09.565'},
     'FastestLap': {'rank': '1',
      'lap': '56',
      'Time': {'time': '1:29.187'},
      'AverageSpeed': {'units': 'kph', 'speed': '214.053'}}}]}

Sample Pandas Output:

	Posittion	RaceName	SeasonYear	RaceNumber	RaceDate	DriverFirstName	DriverLastName	DriverID	StartingGridNumb	ConstructorName
0	1	Australian Grand Prix	2015	1	2015-03-15	Lewis	Hamilton	hamilton	1	Mercedes
1	1	Malaysian Grand Prix	2015	2	2015-03-29	Sebastian	Vettel	vettel	2	Ferrari
2	1	Chinese Grand Prix	2015	3	2015-04-12	Lewis	Hamilton	hamilton	1	Mercedes
3	1	Bahrain Grand Prix	2015	4	2015-04-19	Lewis	Hamilton	hamilton	1	Mercedes
4	1	Spanish Grand Prix	2015	5	2015-05-10	Nico	Rosberg	rosberg	1	Mercedes