ETL Project

**Coast 2 Coast Cuisine**

December 13, 2018

Audrey Bouldin

Nichelle Nesbit

Table of Contents

[Goal 2](#_Toc532456942)

[Objective 2](#_Toc532456943)

[Guidelines/Criteria 2](#_Toc532456944)

[Extracting Data 3](#_Toc532456945)

[Transforming Data 4](#_Toc532456946)

[Loading Data in to Database 10](#_Toc532456947)

[Observations/Findings 10](#_Toc532456948)

[Learnings 10](#_Toc532456949)

[Results 11](#_Toc532456950)

# Goal

Evaluate restaurants in two cities regarding their food inspections grades. Use this data to give recommendations on locations to eat.

# Objective

Perform ETS while using two (2) or more sources of data. Create a final production database and load it into a relational or non-relational source.

# Guidelines/Criteria

Use data food inspection reports for the cities of Los Angeles and New York City.

Use the last 3 full years of data for the analysis.

Locations with “A” ratings have been identified a “excellent” restaurants.

The City of Angels (Los Angeles) is being used as representation of the West Coast cuisine.

The Big Apple (New York City) is being used as representation of the East Coast cuisine.

Any PASSWORD information will be referred to as “PASSWORD” and not shown in writing.

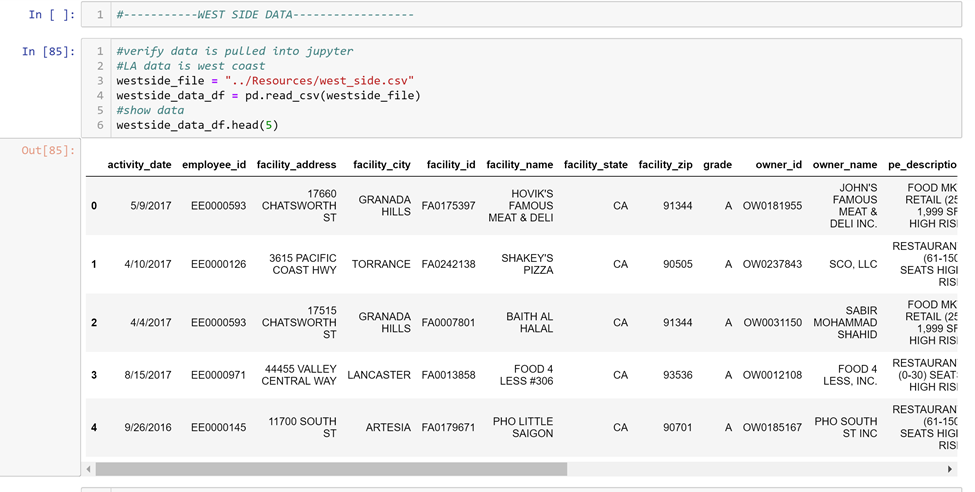
# 

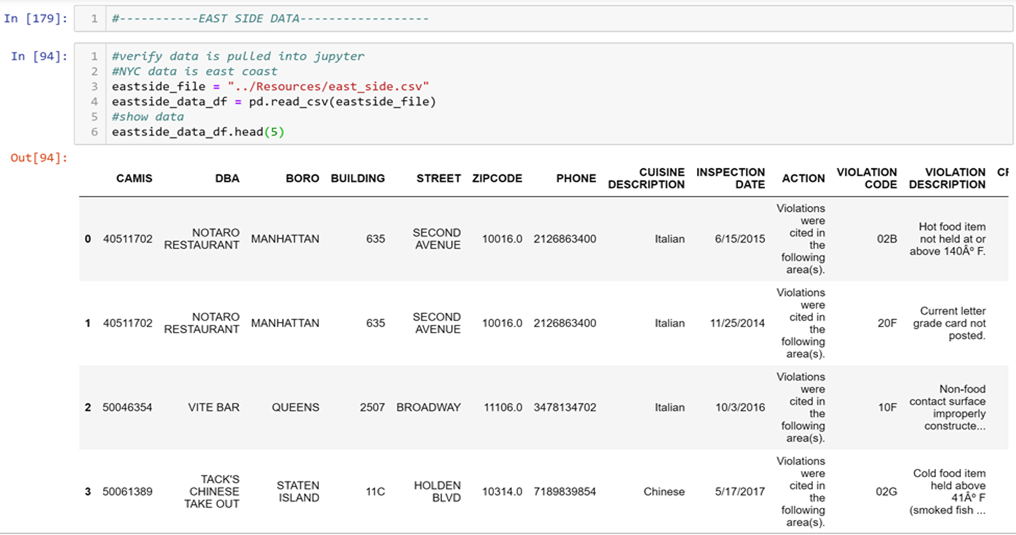
# Extracting Data

Used the following data from Kaggle.com for the analysis:

<https://www.kaggle.com/meganrisdal/la-county-restaurant-inspections-and-violations>

<https://www.kaggle.com/new-york-city/nyc-inspections>

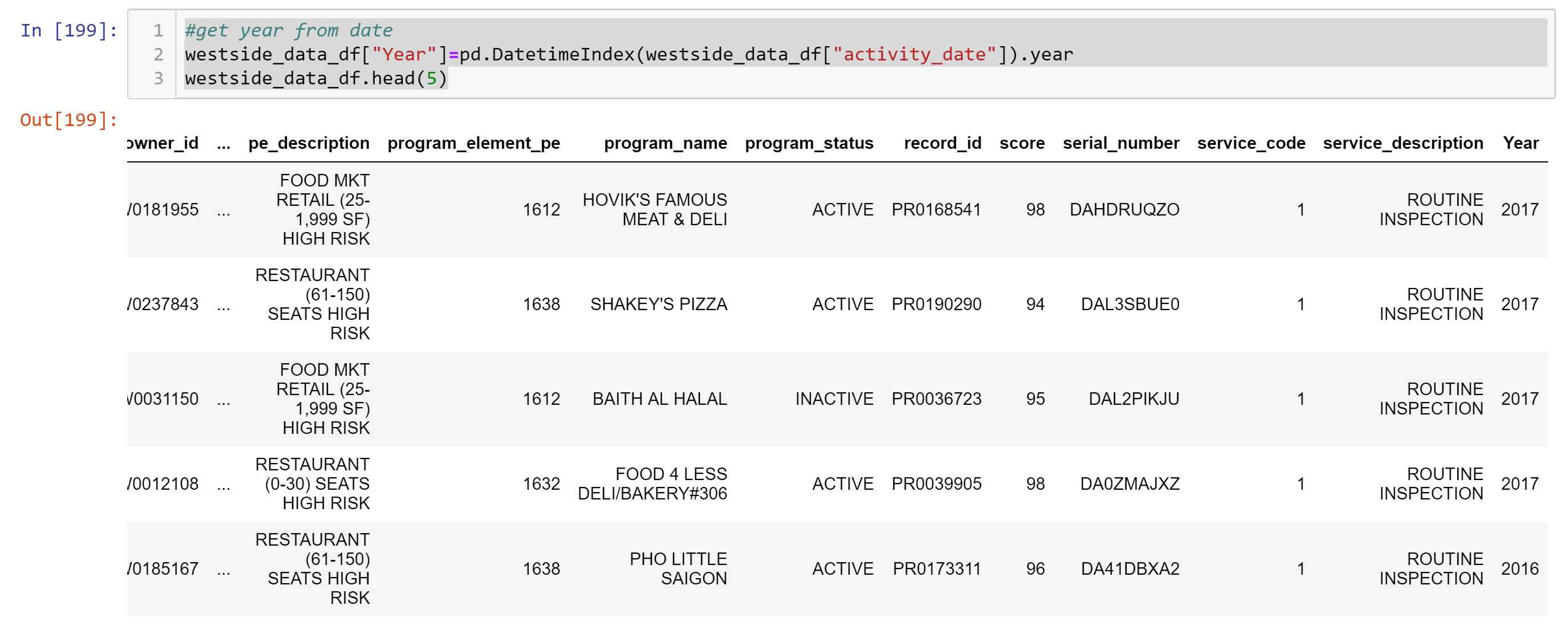


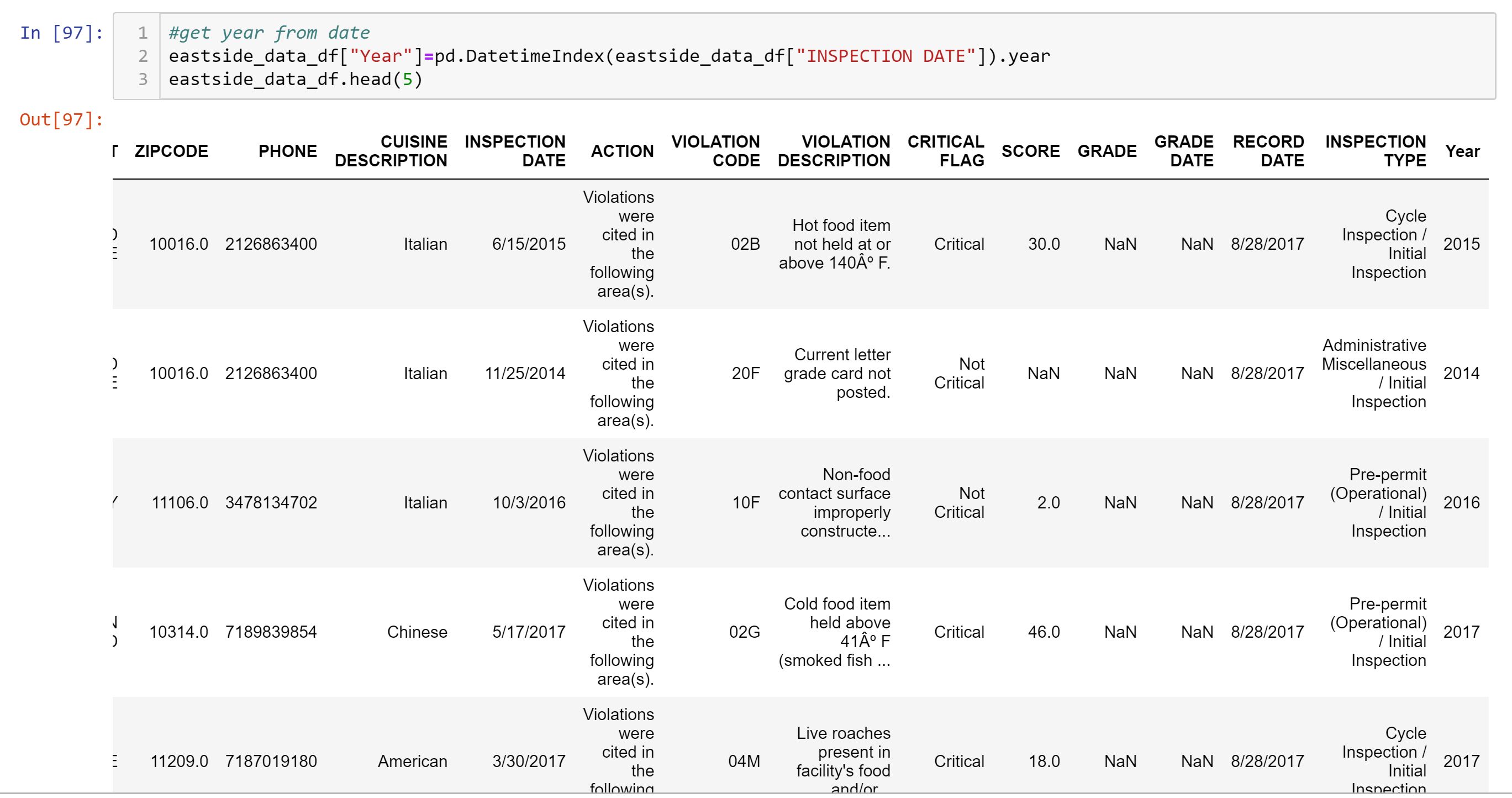


# Transforming Data

**CHANGING DATE:**

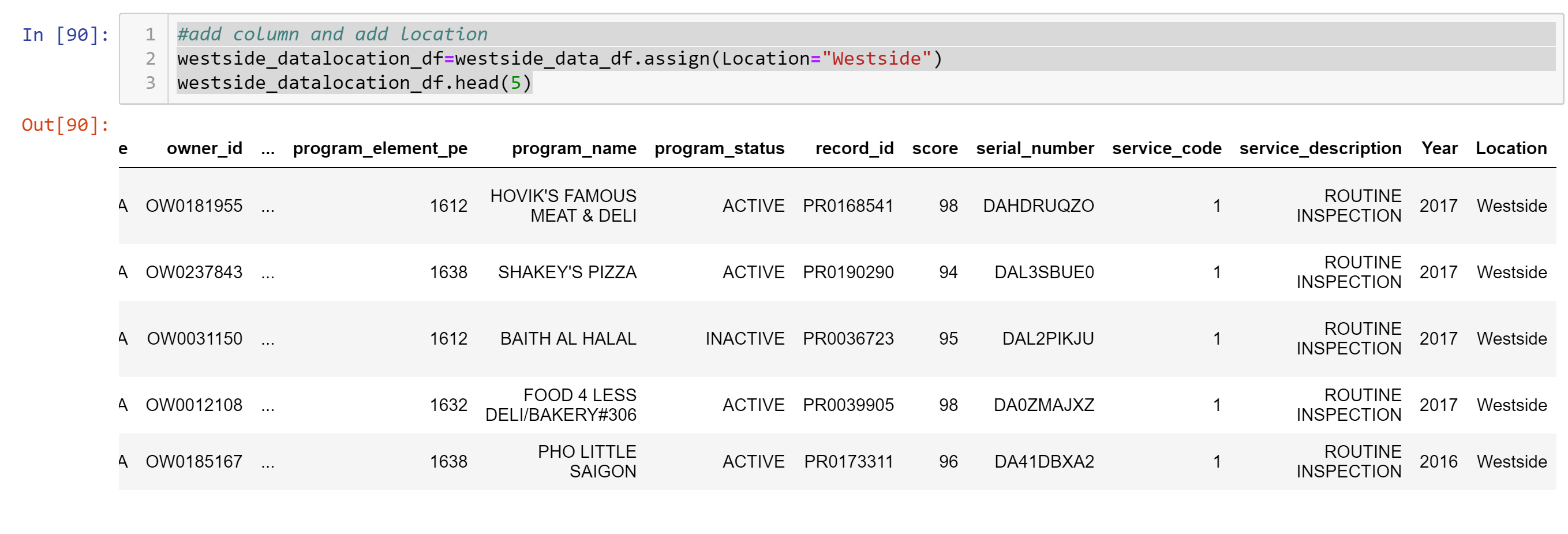
Desired to have data for the last 3 full calendar years. This would include years 2015, 2016, and 2017. See below, data was transformed to add 4 digit year.

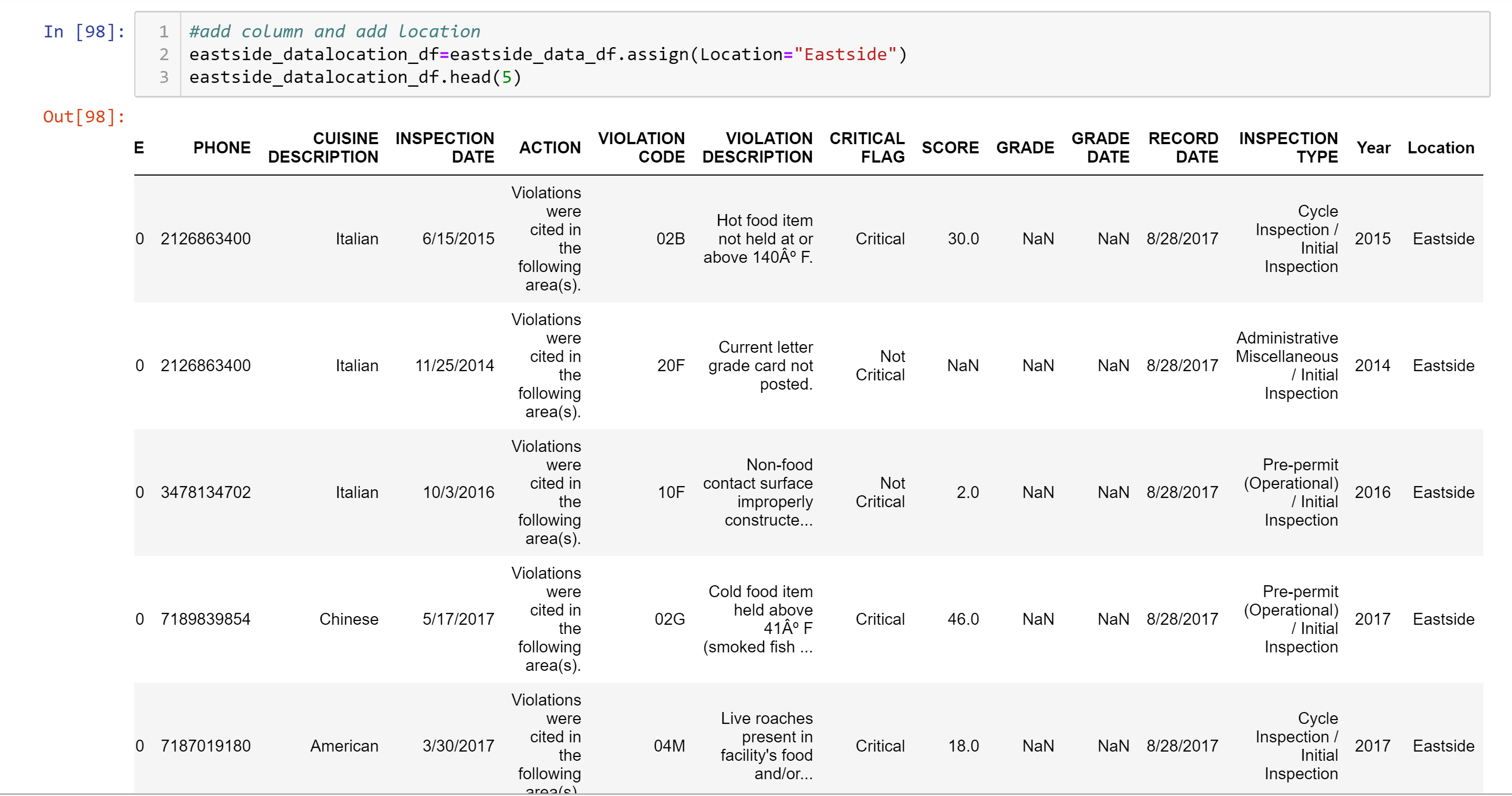




**IDENTIFYING WESTSIDE AND EASTSIDE LOCATIONS:**

See below, location was added based whether it was located on the westside or eastside.





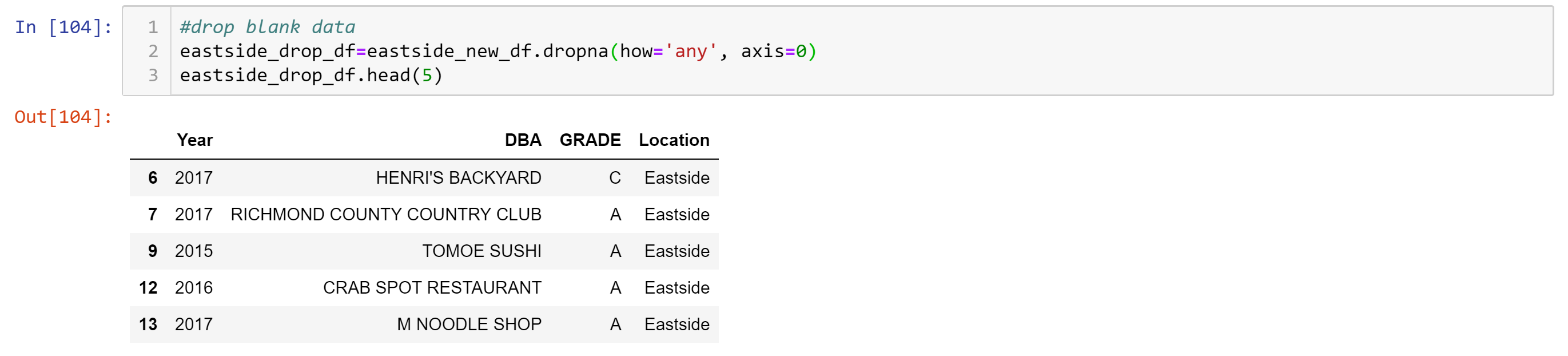
**SELECTING ONLY THE COLUMNS NEEDED FOR THE ANALYSIS:**



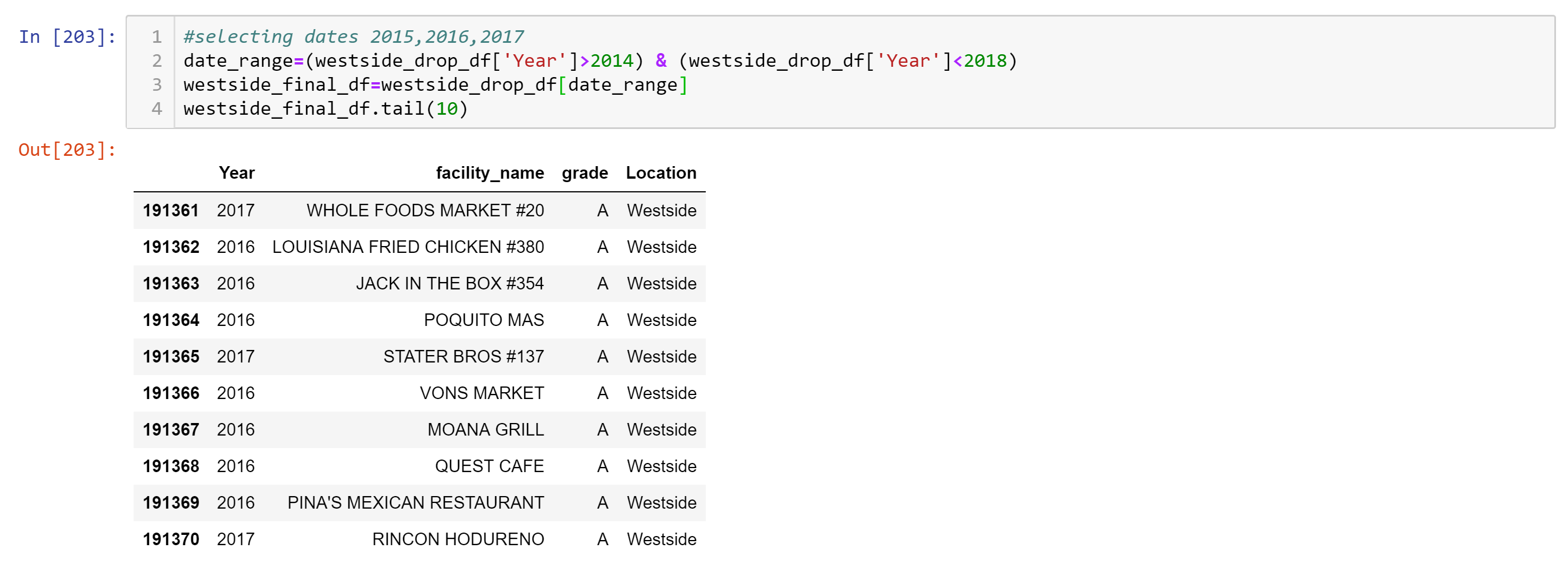


**DROPPING “BLANK” DATA:**





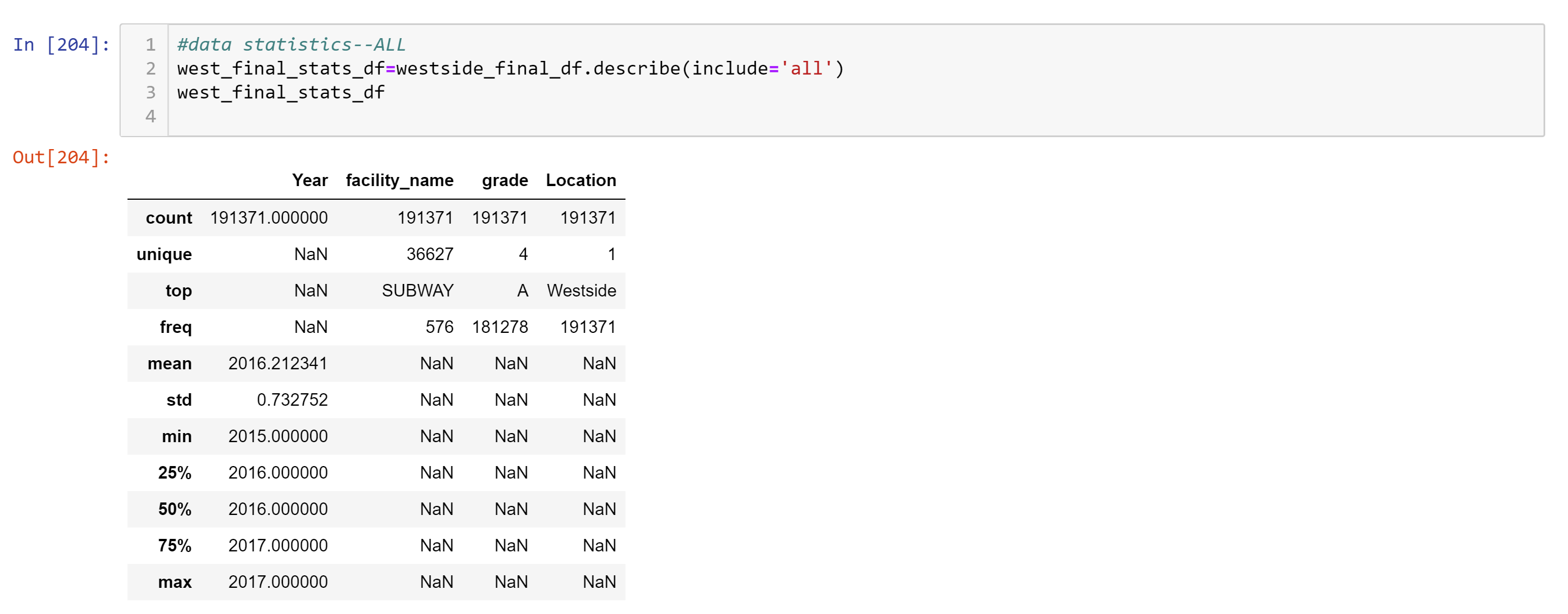
**SELECTING YEARS NEEDED FOR ANALYSIS:**

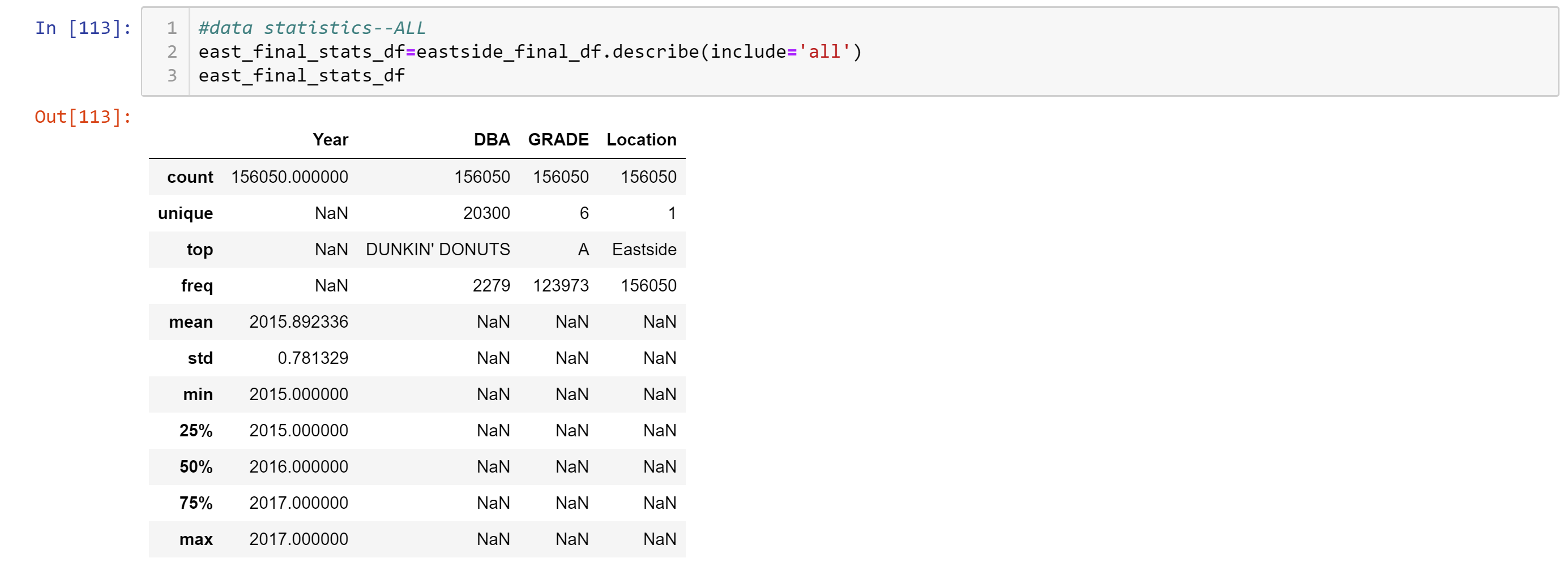


# 

**VALIDATING THAT ALL DATA WAS DROPPED:**

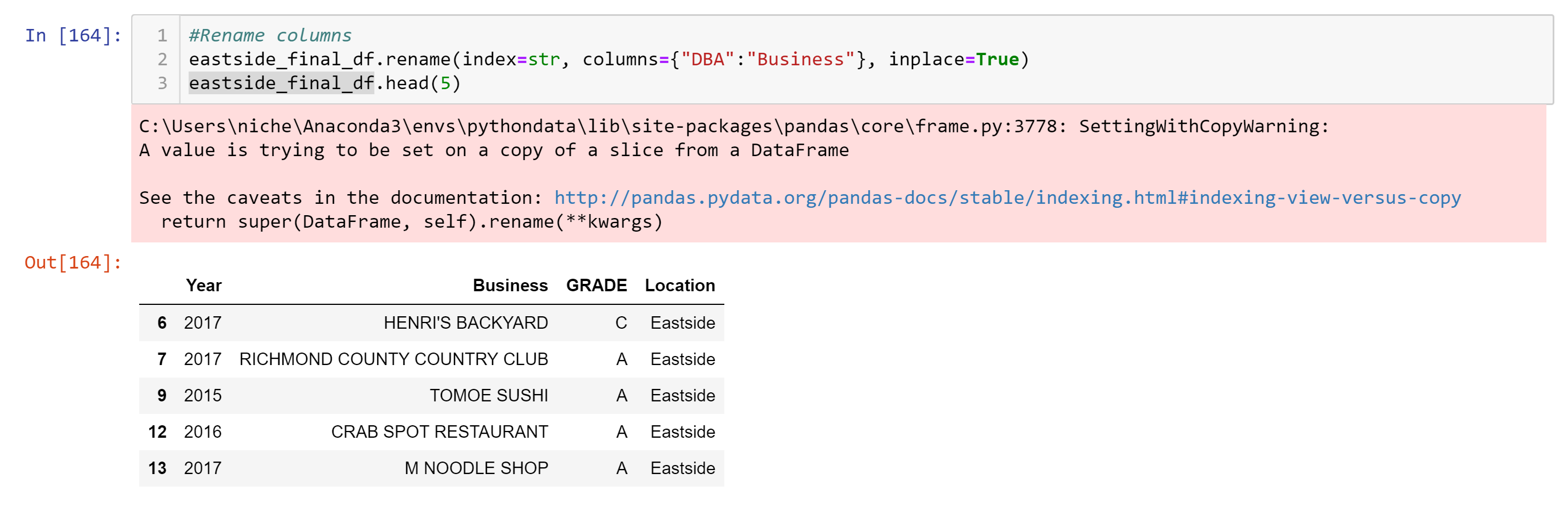
Since the COUNT is the same for all variables, the dropped cells have been removed and we are working complete data sets in regard to each variable.





**RENAMING COLUMNS FOR CONSISTENCY WESTSIDE AND EASTSIDE DATA:**





# 

# Loading Data in to Database

We chose to use a rational database and prepared to load the data into MySQL via the following steps. We loaded two separate tables (westside and eastside).

50 rows of data were selected to make data manageable during the loading into MySQL. This data will be representative of our overall data sets for our analysis.





# Observations/Findings

Westside data comprised of 191371 inspections

Eastside data comprised of 156050 inspections

# Learnings

Revisited rational and non-rational data sets.

Large data sets are difficult to load into other systems (i.e., Github Repository and MySQL Workbench). We had to get creative with the ETL process.

Although we could have added more columns from the original data, to add more scenarios to the MySQL aggregation, the limitation on loading the data was a huge constraint.

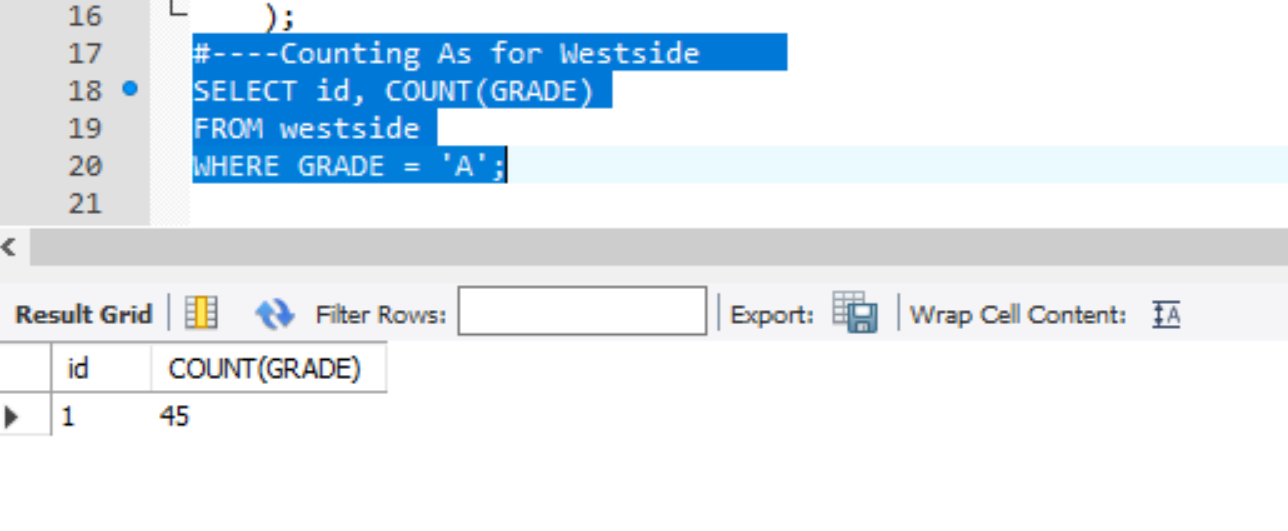
We are truly “hands on learners”. Performing the ETL process was very helpful.

It is important assign data correctly in MySQL. CONVERT was helpful allowing a TEXT to be used for an inequality.

# Results

Based on our samples used and aggregated in MySQL, we recommend the following as “EXCELLENT” restaurants.

We have 39 restaurants with a grade of an A for the Westside.



We have 45 restaurants with a grade of an A for the Eastside.

