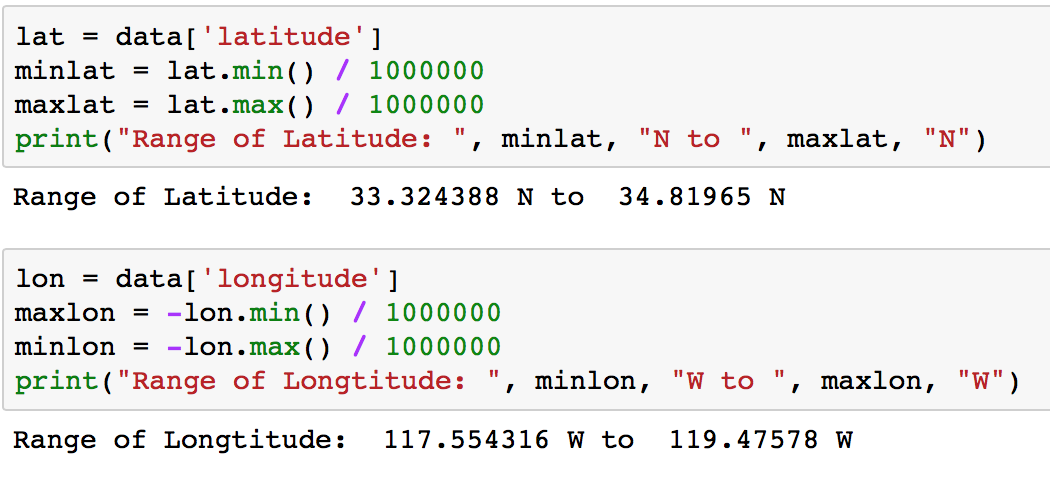
**Report of Data Analysis**

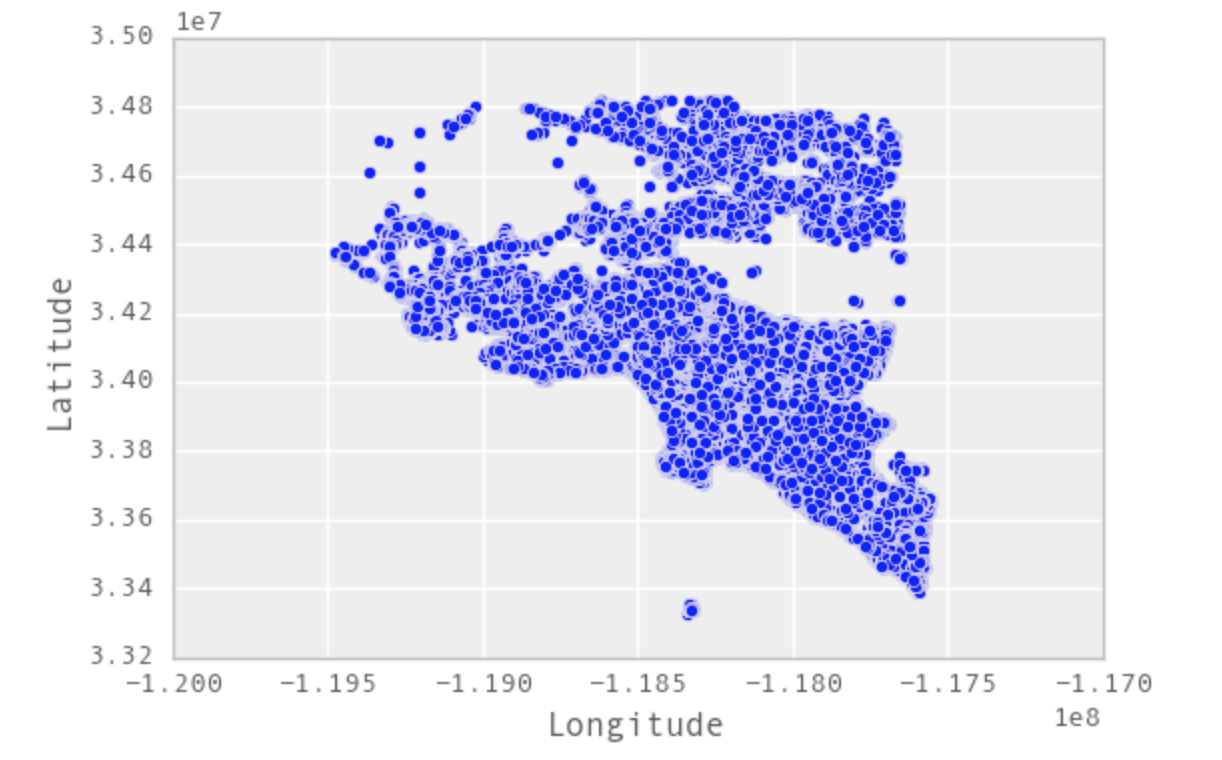
1. **Clean Data**

* Columns that most of values are Nan (most are type column): keep the nan and do not fill with any value, such as ‘hashottuborspa’, ‘heatingorsystemtypeid’;
* Columns that is numerical and presents count: replace nan with number 0, such as ‘fireplacecnt’;
* Columns that is numerical and presents square feet or something else: replace nan with mean of this column, such as ‘finishedsquarefeet12’, ‘lotsizesquarefeet’
* Columns that only has two values (True and False): replace nan with False

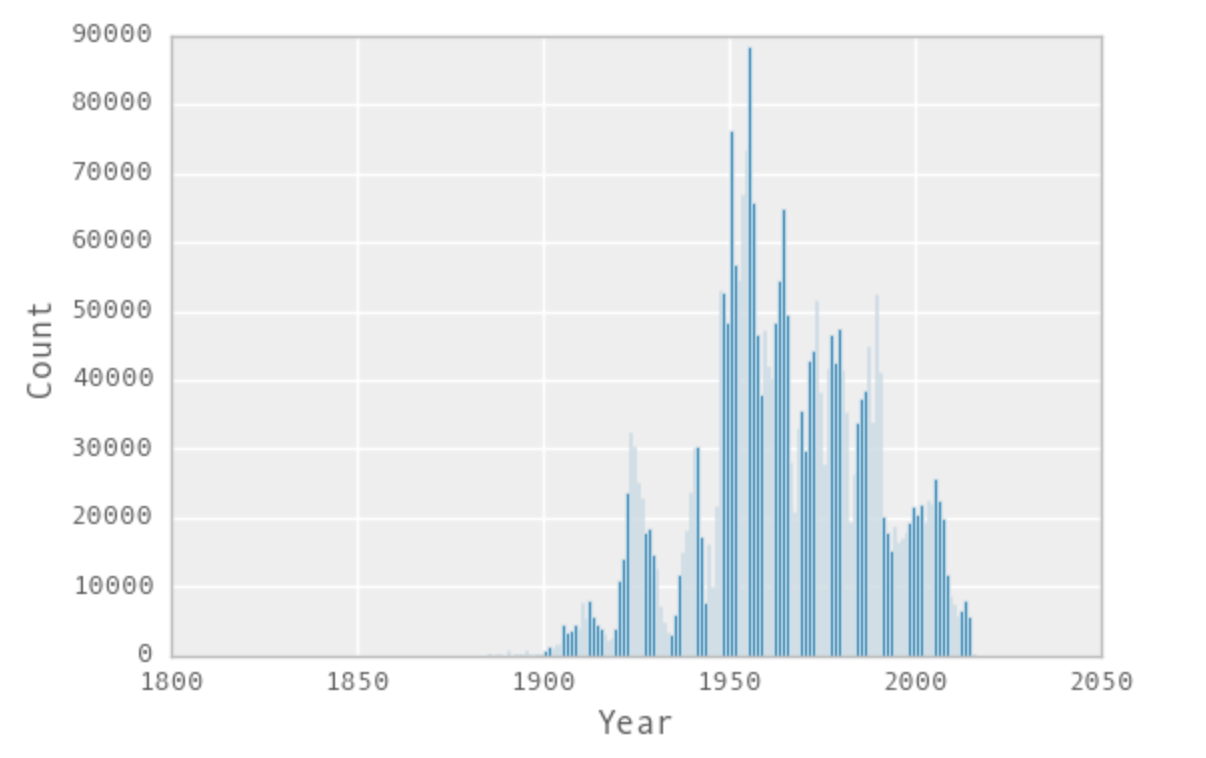
1. **EDA**

* **Calculate the ranges of longitude and latitude, and plot all house location**

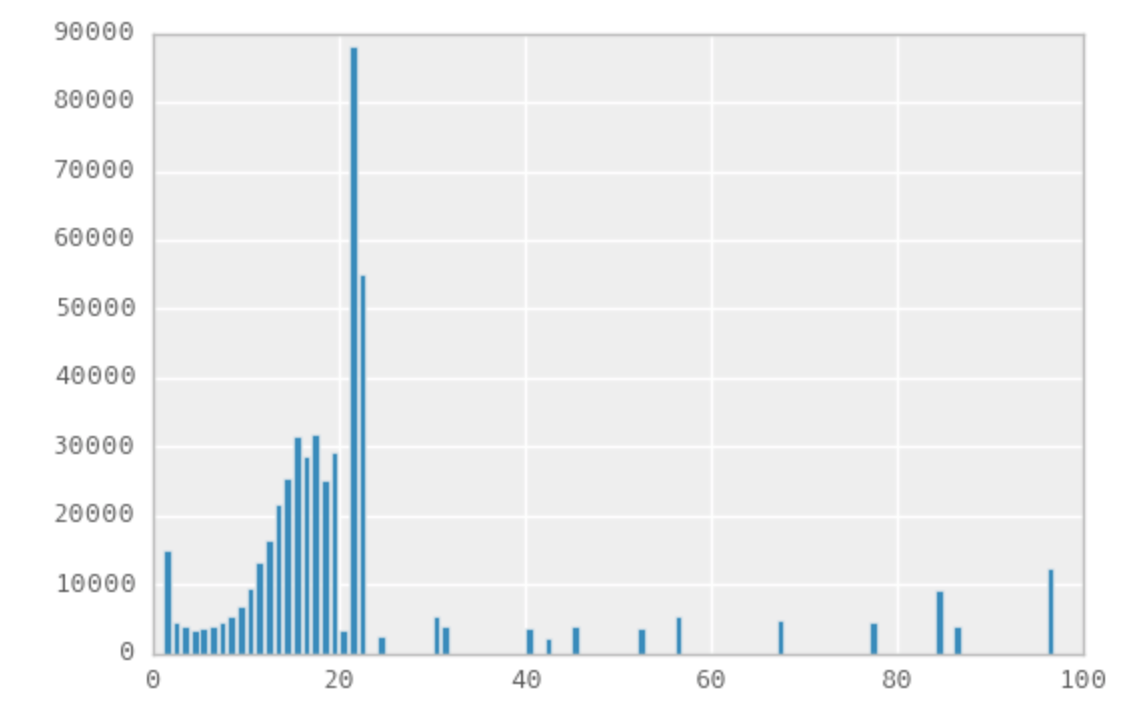
****

****

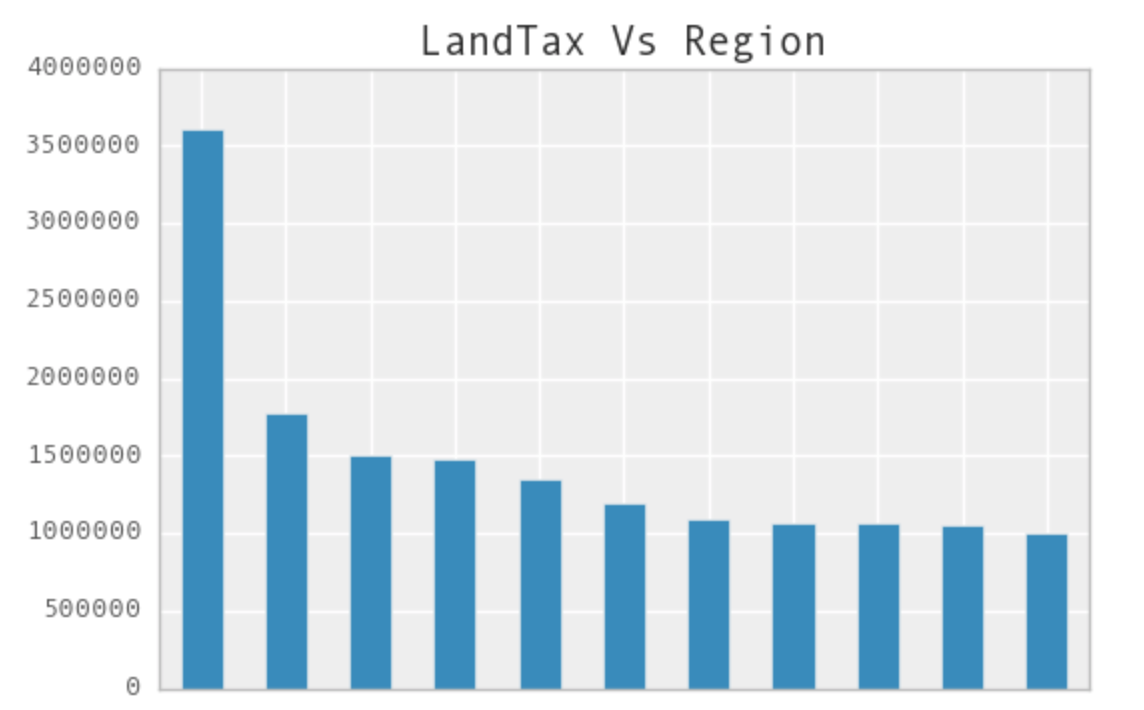
* **House Count Vs Building Year**

****

* **Tax Amount Vs Room Count**

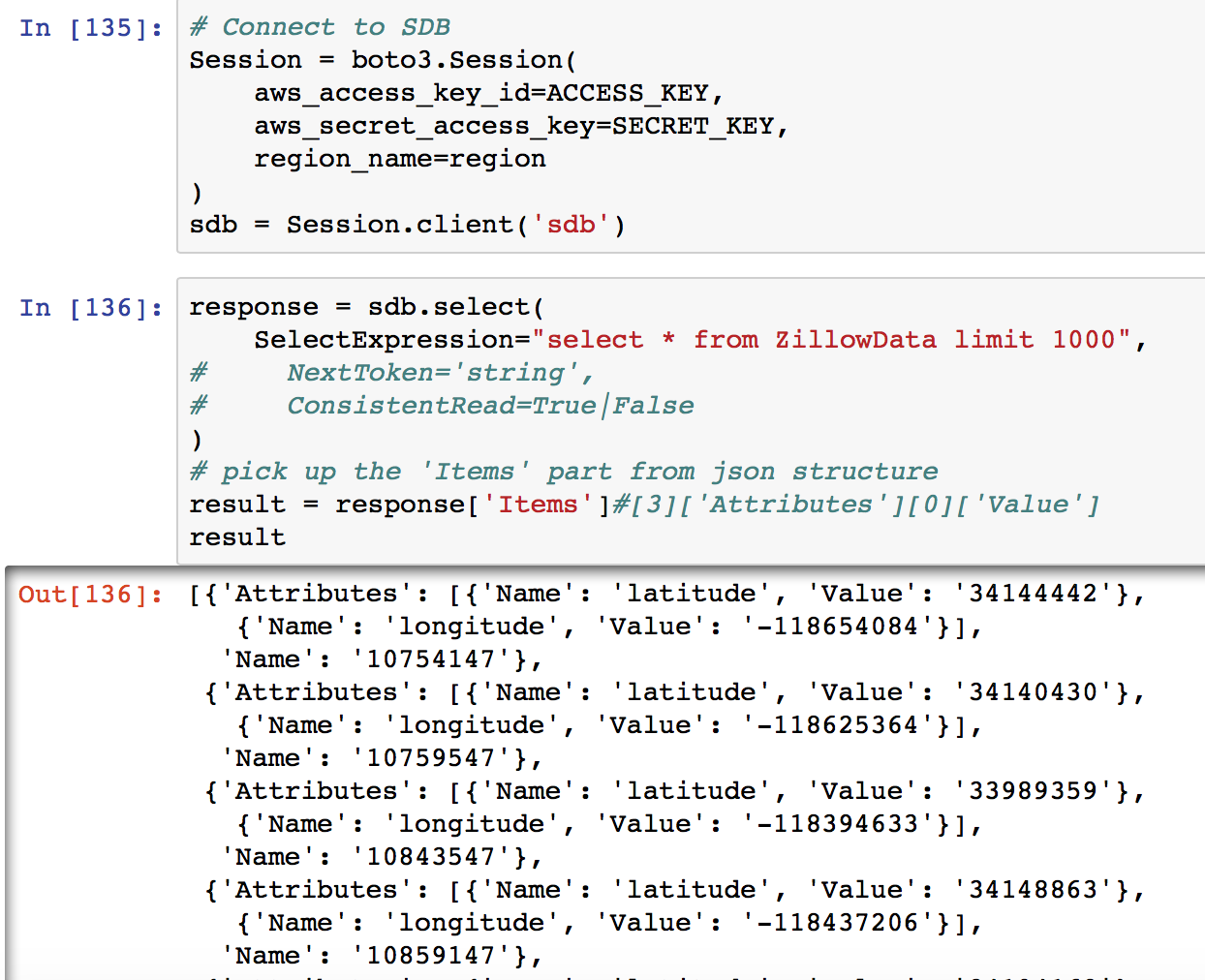
****

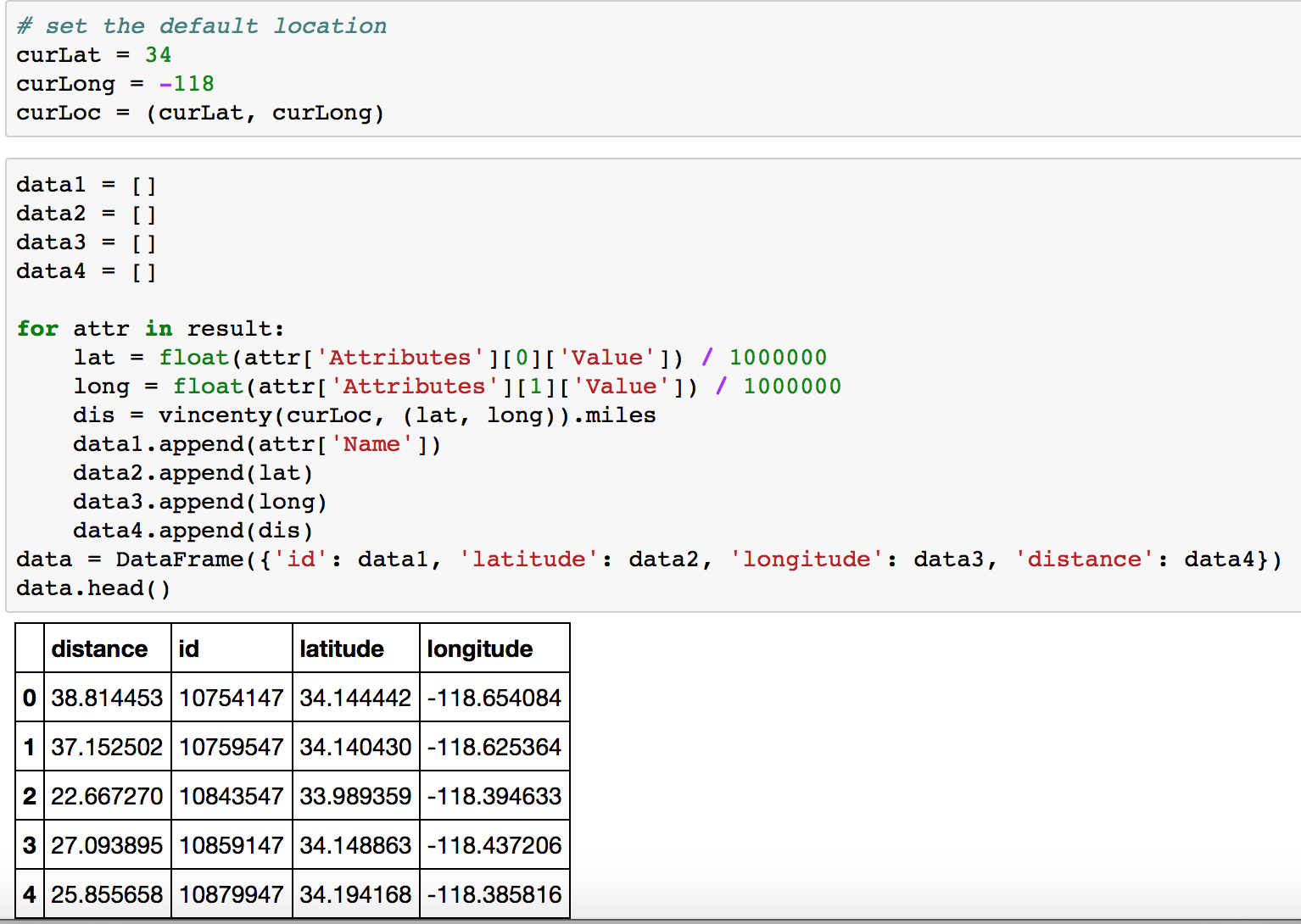
* **Top 10 highest Land Tax Region**

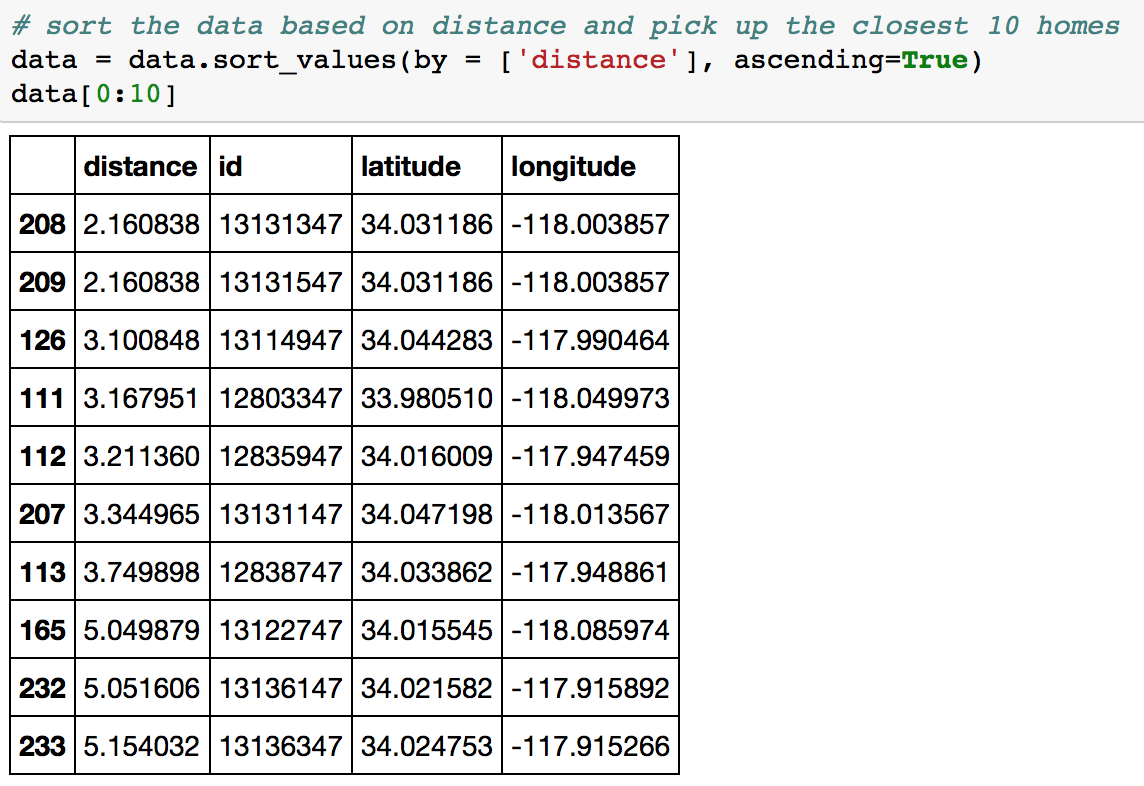
****

1. **Geospatial Search:**

At the beginning, we planned to add conditions in query statements, and get the closest location information. However, our assigned database, Amazon Simple DB, doesn’t support basic operations, so we had to download whole location information, calculated and sorted the distance locally, then picked up the closest 10 location to display in API.



****

****