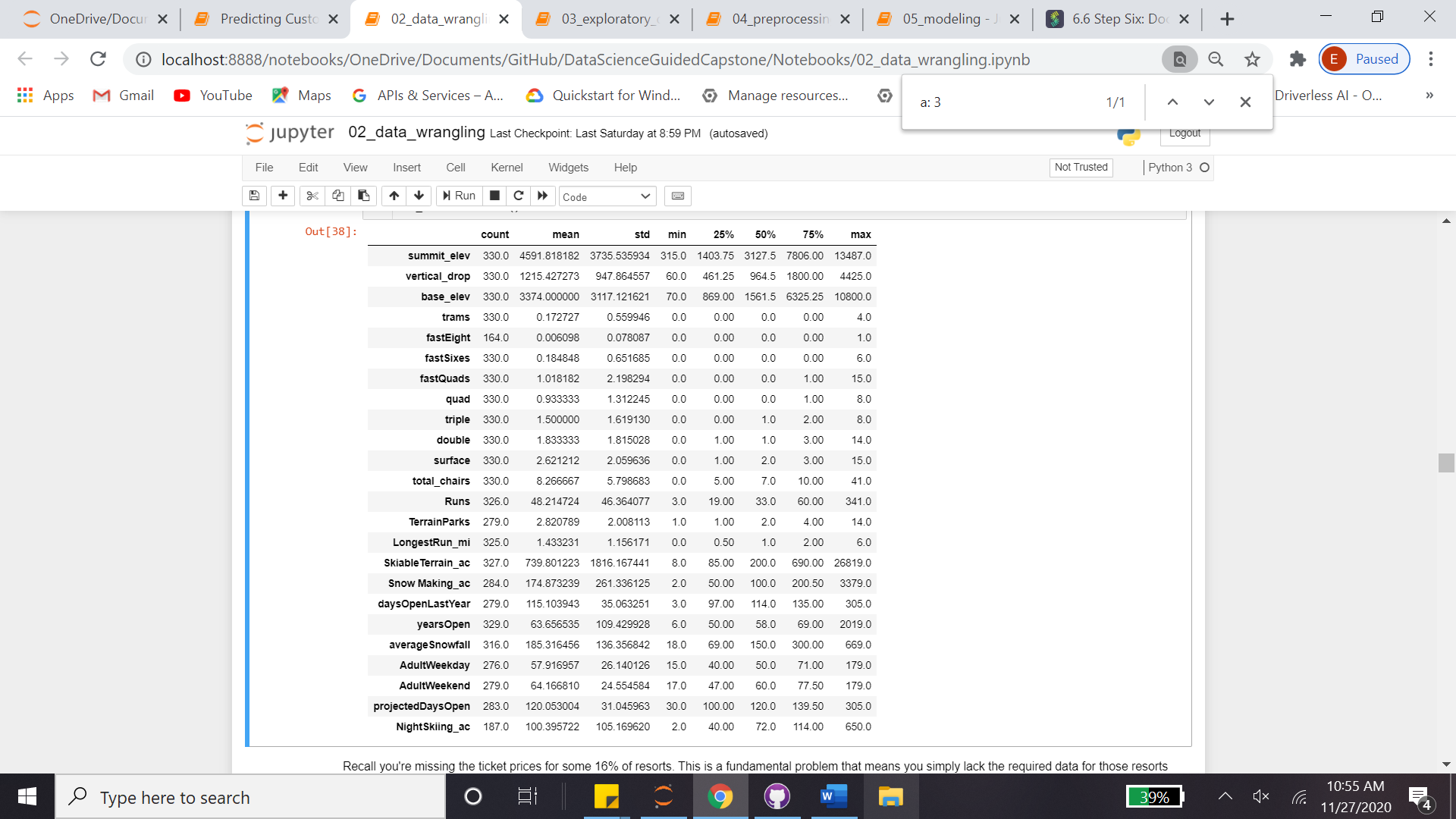
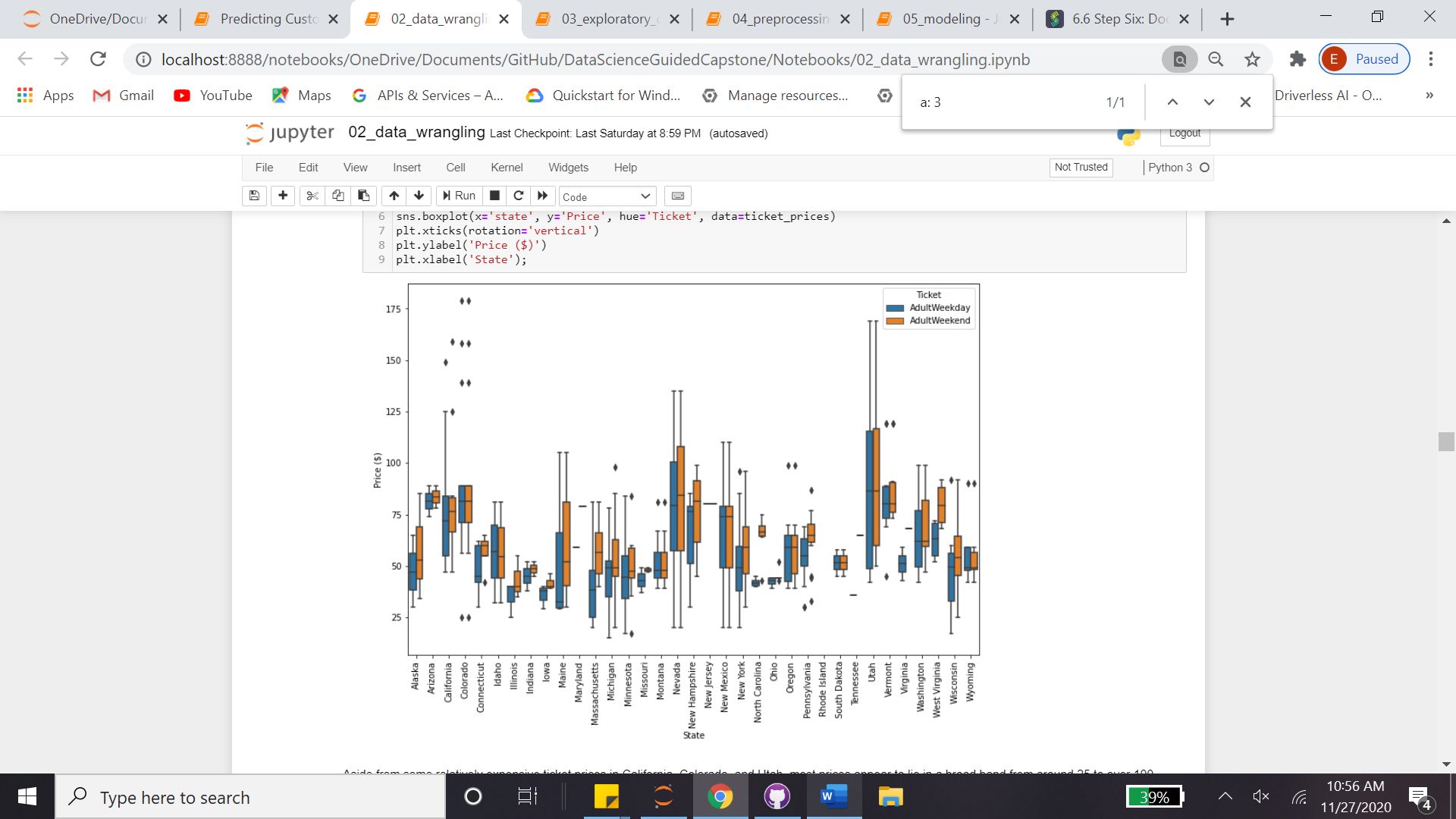
The Ski and Snowboard Resorts industry is expected to continue growing at a higher rate over the five years to 2020. As per capita disposable income continues to increase, many Americans will likely continue spending on vacations and activities, such as skiing and snowboarding. IBISWorld anticipates industry revenue to increase at an annualized rate of 3.6% to $3.4 billion over the five years to 2020.

Big Mountain suspects it may not be maximizing its returns, relative to its position in the market. It also does not have a strong sense of what facilities matter most to visitors, particularly which ones they're most likely to pay more for. This project aims to build a predictive model for ticket price based on a number of facilities, or properties, boasted by resorts (at the resorts).

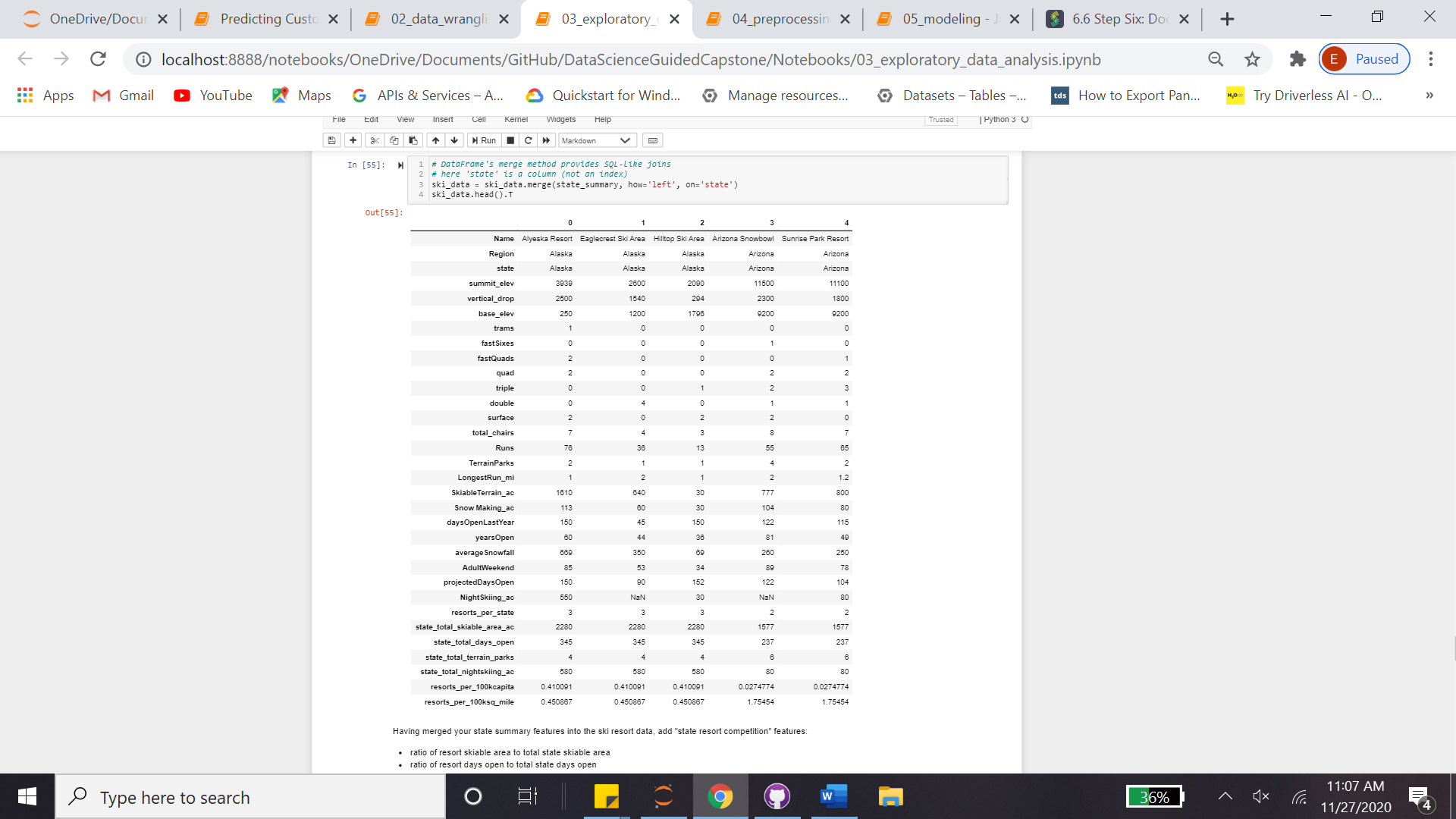


Two columns were dropped, namely 'fastEight' and 'AdultWeekday'. Reasons were non-informative (too many zero) and repetitive (similarity to 'AdultWeekend') data respectively.

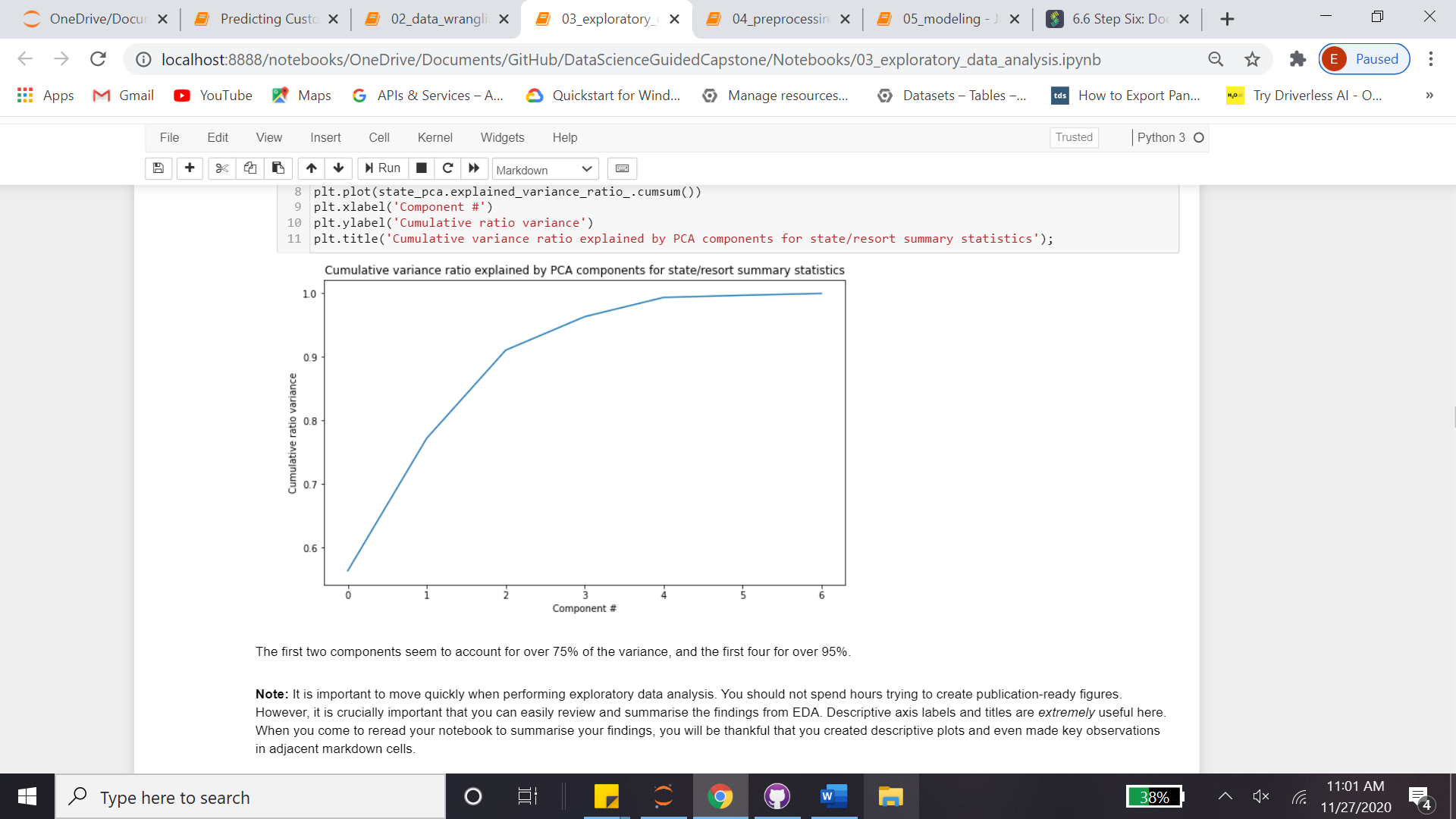


**Feature discussions**

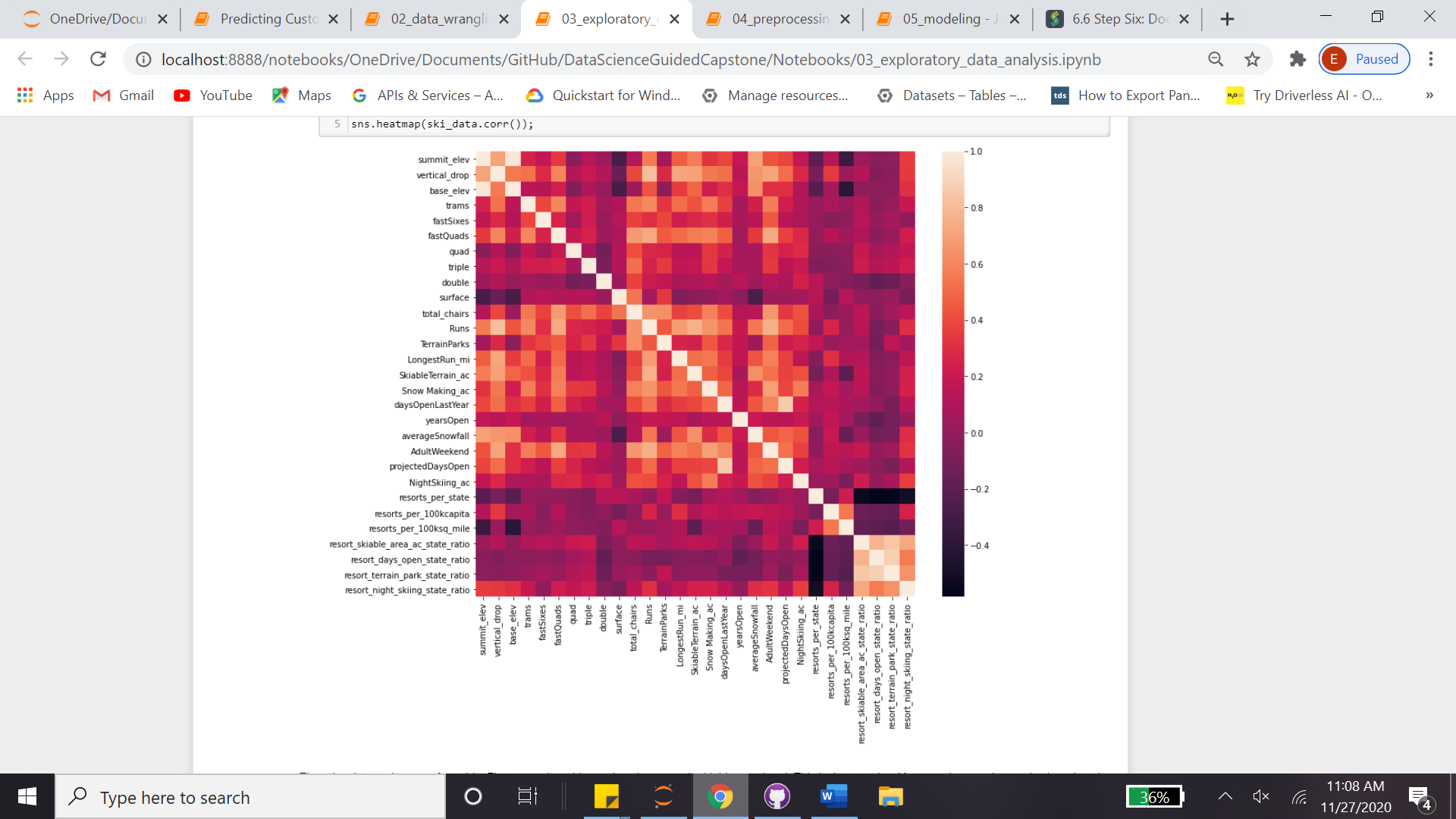
Target column was identify as 'AdultWeekday', aka 'Price'. Features were engineered per four columns: TerrainParks, SkiableTerrain\_ac, daysOpenLastYear and NightSkiing\_ac. Per Wikipedia demographics, they were aggregated per sum and added as new columns



6 out of 7 newly created features dominate 75% of the variance, to see the correlation two dataset were merged into one

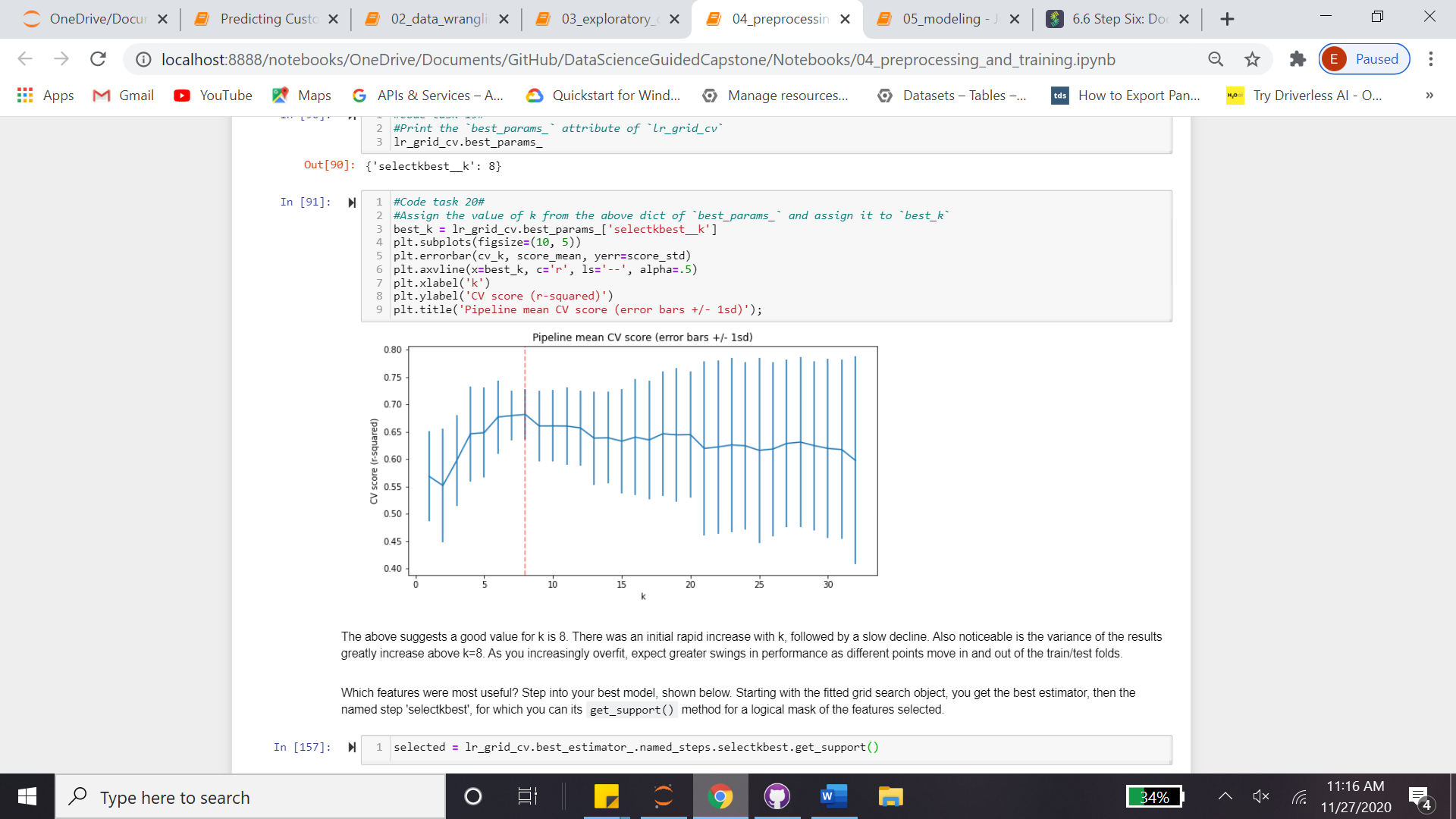


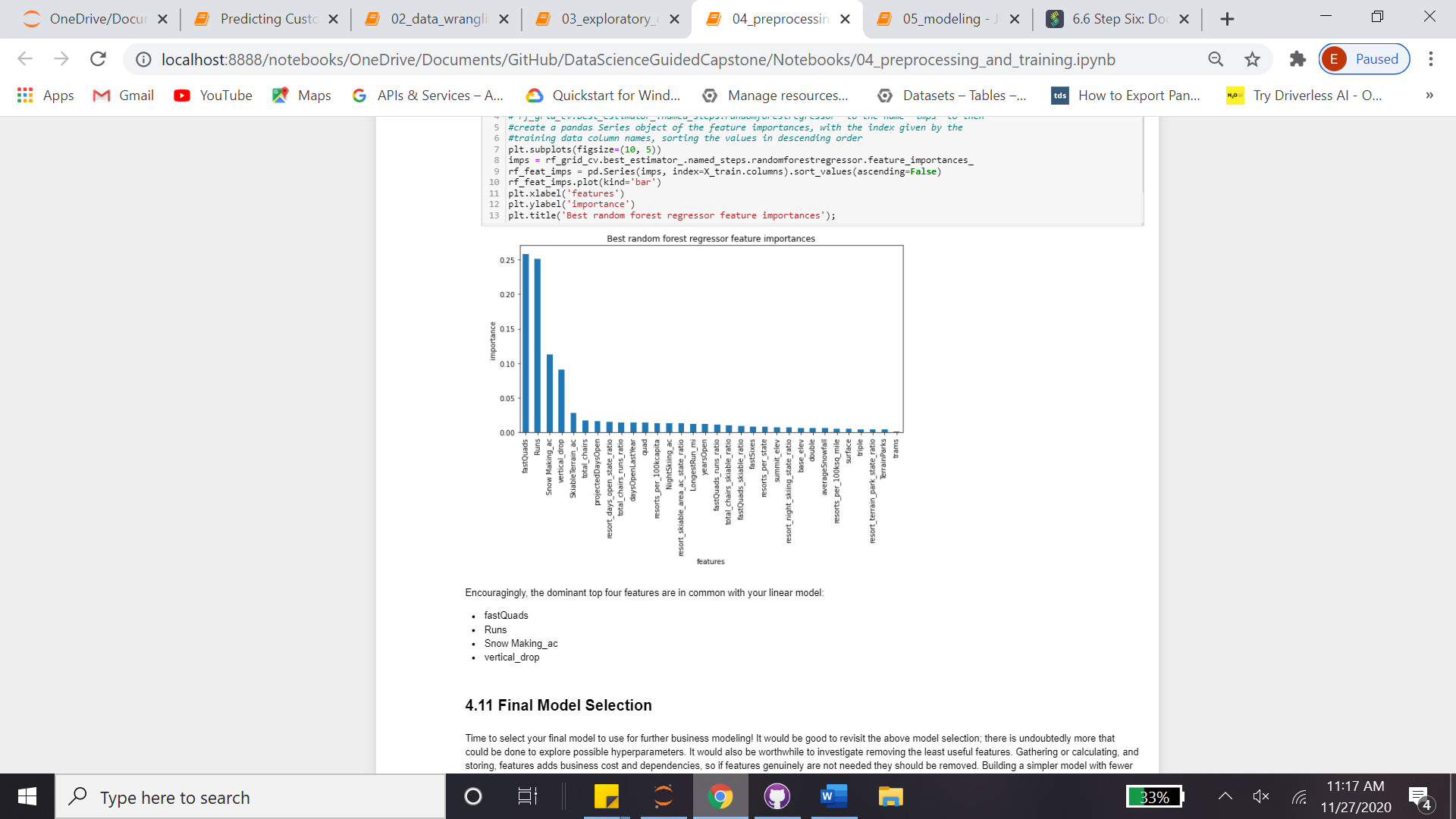
Most significant correlation were observed with fastQuads, Runs and Snow Making\_ac. The last one is interesting. Visitors would seem to value more guaranteed snow, which would cost in terms of snow making equipment, which would drive prices and costs up.



**Fitting & Predicting**

Of the new features, resort\_night\_skiing\_state\_ratio seems to have mild correlation. This also might be considered as a subject for incremental investment. First, R2 value is examined per mean of the target, 0.7-to-0.8 looks promising. Comparison by both median and mean were applied and the former favored a bit more. The whole data is scaled by StandardScaler, imputed with median, as well as mean. Train/Test split as 70%/30% (Later on, it's validated by learning\_curve). Pipes were created for both linear regression and random regressor, where the latter is performed better. CV was significantly varying, best\_params was applied and gave only 8 features. Random regressor parameters were reset, and iterated again, the final MAE is noted as 9.54, where the mean is 63.81





**Suggested improvements**

* There are two gaining possibilities, either predicting a price increase/decrease or cost saving by feature reducing: The best model was applied and it observed that Big Mountain Resort modelled price is 95.87 USD, where actual price is 81.00 USD. Even with the expected mean absolute error of 10.39 USD, this suggests there is room for an increase. 5 USD increase may sum a total of 7.5M USD (350K guests per 5 days stay in average) per year.
* Per cost saving, several change proposals were noted.
  + Adding a run, increasing the vertical drop by 150 feet, and installing an additional chair lift; 1.99 USD increase that may sum 3.5M USD per year. (Adding 2 acres of snow making on top of this does not look like having incremental effect.)
  + Close up to 10 of the least used runs -> Very slight price decrease with up to 6 (excluding) close ups, which makes it applicable.

**Future directions**

* For instance, closing runs may have other benefits like decreasing the costs, other than our primary target which is Ticket Price.
* Second, when we increase the lenght of the run, we're having some gain with the pricing, however this is not a net value, since we don't know the size of the investement we need for the extra chairs.
* This scenario calls for increasing the longest run by .2 miles and guaranteeing its snow coverage by adding 4 acres of snow making capability makes no difference, 0 USD