

Python Data Science Test

Summary

In order to test your ability to use Python and other tools to understand/ manipulate data, we have created a few simple tasks to complete. Feel free to use other tools such as Excel and Tableau, but the majority of work should be done with Python. Also, feel free to Google any terms that you don't understand.

The Test

- 1) Make sure you have the file "re_data.csv"
- 2) Use Python to answer these questions:
 - a. Which state has the highest average loan amount?
 - b. Which state has the highest taxes as a % of property value?
 - c. Which variable is the strongest predictor of maintenance expense (as a percentage of property value)? What is your interpretation of the result?
 - d. What is the median, range and variance in Loan to Value of our loans? What might this tell us about how deterministic loan amount calculations are?
 - e. Do properties with larger valuations exhibit any economies of scale with regard to expenses? Why do you think this is?
- 3) Please create graphs depicting the following. Feel free to use any tools that you think would help.
 - a. Expense ratio (Total Operating Expenses/ Effective Gross Income) vs state
 - b. Relationship between age of property and expenses
 - c. Cumulative loan amount by fiscal quarter
- 4) Using Scikit-Learn or a library of your choice, create a model that predicts loan to value. How accurate is your model? What would be required to improve this score?
- 5) Commit the code you used to GitHub/ Bitbucket.
- 6) Send the GitHub URL and graphs to robert.grzesik@greyco.com and zachariah.rosenberg@greyco.com as soon as you are finished.