Logo

Description automatically generated

AML-2203 Advanced Python AI and ML Tools

Assignment

(20% of Final grade)

|  |  |
| --- | --- |
| **Instructor:** | Vahid Hadavi, PhD, P.Eng |
| **Class:** | **AML-2203** |

# Description

This assignment requires more individual learning than previous activities - you are encouraged to check out the [pandas documentation](http://pandas.pydata.org/pandas-docs/stable/) to find functions or methods you might not have used yet, or ask questions on [Stack Overflow](http://stackoverflow.com/) and tag them as pandas and python related. And of course, the discussion forums are open for interaction with your peers and the course staff.

Definitions:

* A quarter is a specific three month period, Q1 is January through March, Q2 is April through June, Q3 is July through September, Q4 is October through December.
* A recession is defined as starting with two consecutive quarters of GDP decline, and ending with two consecutive quarters of GDP growth.
* A recession bottom is the quarter within a recession which had the lowest GDP.
* A university town is a city which has a high percentage of university students compared to the total population of the city.

**Hypothesis**: University towns have their mean housing prices less effected by recessions.Perfome the following tasks:

* Download the data and run a completed set of exploratory data analysis including best possible tasks on that. This may include but not limited to observing the shape, describe the data set, checking the missing values, drawing the profile of the data, checking the distribution type and data types and several other tasks.
* Run a t-test to compare the ratio of the mean price of houses in university towns the quarter before the recession starts compared to the recession bottom. (price\_ratio=quarter\_before\_recession/recession\_bottom)

The following data files are available for this assignment:

* From the [Zillow research data site](http://www.zillow.com/research/data/) (https://www.zillow.com/research/data/) there is housing data for the United States. In particular the datafile for [all homes at a city level](http://files.zillowstatic.com/research/public/City/City_Zhvi_AllHomes.csv), City\_Zhvi\_AllHomes.csv, has median home sale prices at a fine grained level.
* From the Wikipedia page on college towns is a list of [university towns in the United States](https://en.wikipedia.org/wiki/List_of_college_towns#College_towns_in_the_United_States) which has been copy and pasted into the file university\_towns.txt.
* From Bureau of Economic Analysis, US Department of Commerce, the [GDP over time](http://www.bea.gov/national/index.htm#gdp) of the United States in current dollars (use the chained value in 2009 dollars), in quarterly intervals, in the file gdplev.xls. For this assignment, only look at GDP data from the first quarter of 2000 onward.

Your report must include your Jupyter notebook including the detail of your code and proper markdowns and a pdf file explaining your approach and process, results and what your learned in this activity.