University of California Life Sciences Graduate Programs and the Role of Housing Costs Near Campus

[redacted]

Western Governors University

Table of Contents

A. Proposal Overview

A.1 Research Question or Organizational Need

Are doctoral graduate program applications for life sciences programs on a declining trend at the University of California? Does the cost of housing for students near UC campuses correlate with application characteristics such as application or admission rates?

A.2 Context and Background

Some students and faculty in the neuroscience department at a University of California campus fear that the incoming cohorts of graduate students seem to be smaller than in previous years. The University of California is a research institution – that is to say the research performed is the fundamental source and basis to its students, their programs, and all business performed at the University.

Graduate students are the workforce behind research being performed – the basic necessity to funding for research laboratories, and the university itself. If fewer students are entering the field, some research groups could find difficulty performing research necessary for continued funding. New students also learn a great proportion of their techniques from senior students. When large enough gaps appear in the training of new students, many techniques can be lost, creating great obstacles in the continued effectiveness of laboratories across the university.

Students are also facing great difficulty affording housing near campus. Graduate students are responsible for research work that requires irregular hours and on-call availability. Almost always, graduate students are unable to commute, and must live in the near surrounding area around their campus. Since graduate students have very limited funding, the cost of housing near campus greatly affects their ability to perform research and attend these very long programs.

A.3 and A3A Summary of Published Works and Their Relation to the Project

Review of Work 1

The Academic Impact of Financial Stress on College Students"

This paper, published in Sage Journals, highlights the correlation between student financial stress and academic success. They state that, "Staying in school and graduating on time is an important factor for students and their families. Greater financial burdens may lead students to reduce coursework or drop out of school for paid work." (Joo et al., 2008)

Review of Work 2

"Methodology: Zillow Observed Rent Index (ZORI)"
This methodology article details the Zillow Observed Rent Index, or ZORI. The ZORI measures changes in rent costs over time, while controlling for changes in rental availability. A major issue with simply looking at changes in asking rent over time, is that this method does not account for the changes of rental availability at any given time. If an area were to experience a period of greater or lower availability, average rental costs would be skewed by variance in types of available units at different times. ZORI addresses this issue by aggregating differences in price for the same unit over time. This means listings that appear often are weighed less than listings that appear infrequently over longer periods of time. (Clark, 2022)

Review of Work 3

"Increases in Graduate Student Debt in the US: 2000 to 2016"

Here is study on the increase and related effect of debt in graduate students. This study emphasizes the negative effect on success that increasing debt has on graduate students. Here they make the statement**,** **"…**increasing graduate level debt may curtail other life choices, may discourage students from enrolling and persisting, or may motivate degree earners to pursue different program or career options due to accumulated loans." (Webber & Burns, 2020)

A.4 Summary of Data Analytics Solution

In this project I propose to investigate possible correlation between the cost of housing around UC campuses and the respective campus' graduate admission rate characteristics. This will help the administrative staff of the University of California better understand the role of financial housing assistance and student housing opportunities in the success of graduate programs, and more specifically in this case, life sciences doctoral graduate programs.

The first product of this undertaking will be a figures that visualize trends in program application characteristics such as application rates and admissions, accompanied by an associated table of statistics. Secondly this project will produce a visualization and associated table of statistic outcomes representing correlation between housing costs and campus application rates.

The data used for this will be publicly available data from the University of California that includes graduate student information (this will be preprocessed in excel), and publicly available data from Zillow that includes information on the adjusted cost of rent for each campus' home city. The analysis will be performed primarily in python, and some data cleaning will be performed in Excel.

A.5 Benefits and Support of Decision-Making Process

A deeper understanding of the connection between housing cost and graduate program admission directly benefits the decision making of administrative staff at the University of California. Many financial decisions are made regarding the student and family housing availability. Staff must also understand the dynamicism of housing costs in California in order to properly budget stipend categories for students.

B. Data Analytics Project Plan

B.1 Goals, Objectives, and Deliverables

The goal of this project is to provide insight on the role of housing cost near UC campuses in the success of graduate programs and their students measured by their applications characteristics. I will specifically be investigating the life sciences doctoral graduate programs.

The primary objective is to investigate the existence of a correlation between life science graduate program admission characteristics (applications, admissions) and the cost of housing (specifically adjusted cost of rent) near UC campuses. To complete this objective, this project will first deliver a statistic analysis of trends in application characteristics for each campus, utilizing a series of Mann-Kendall tests to determine the significance of these trends. This analysis will include visualizations of each campus' trends as well as a table of Kendal Tau values and associated p-values.

The next objective is to investigate the correlation between the cost of rent near a campus, and that campus' previously determined program application trends. The deliverables required for this objective include tables containing the outcomes of a series of Pearson correlation calculations between rent costs and application numbers, as well as a visualization of the significant outcomes.

B.2 Scope of Project

The scope of this project includes:

- The collection and processing of data from the UC public database, and from Zillow's rental database.
- The analysis of admission characteristic data and rental data to produce relevant summary statistics and statistic visualizations
- Discussion of the delivered statistic analysis

Some considerations outside the scope of this project include non-rental housing costs such as home value or living expenses, as well as considerations of data outside the date range of the data (2015-2022). Also, the deliverables of this project show *correlations* between variables. It is important to remember that *causal* relationships are not being investigated by this analysis and while we can speculate on the cause of correlations, further investigation is outside the scope of this project.

B.3 Standard Methodology

An agile methodology will be applied to the completion of this project. This project has defined goals to be completed by a single individual (myself), and is therefore best suited for a flexible goal oriented methodology. The four foundations of an agile methodology include adaptivity, goal-oreinted structure, integrated approach, and learnable nature. Seeing as this is a single-person project, adaptability is key to this project's success. As the data is investigated, different techniques and programmatic approaches will have to be considered and implemented as their needs arise and become apparent. As stated previously, this project has clearly defined goals – the statstic analysis of defined data sets. The outcome of this project can be directly integrated into the decisions of relevant committees, such as financial committees on campus. Lastly, this project is only the first step of a

series of highly relevant questions concerning graduate programs and student finances. The outcomes of this project should elucidate the path for learning more about these important interactions by revealing deeper, more nuanced questions concerning the role of student finances and academic success.

B.4 Timeline and Milestones

Milestone or deliverable	Duration (hours or days)	Projected start date	Anticipated end date
Create Timeline and Project Plan	7 days	*10/11/2023*	*10/21/2023*
Gather and Clean Data	3 days	*10/21/2023*	*10/24/2023*
Foundational code and analysis of data	10 days	*10/24/2023*	*11/04/2023*
Visualization and Discussion of Statistical Analysis	2 days	*11/04/2023*	*11/06/2023*
Testing/Verification/Revision	1 day	*11/06/2023*	*11/07/2023*

B.5 Resources and Costs

1. Labor Costs: No Cost
2. Software/IDE: No Cost
3. Access Fees: No Cost

All work will be performed by myself for academic purposes, and therefore incur no financial cost. I will be using Python IDE and Jupyter Notebooks which are free to use, I will use my student account to access Microsoft office since I will be using Excel for some data processing, and I will access only public data that has no access fee.

B.6 Criteria for Success

The success criteria for this project are based upon the completion of its deliverables. To be considered a complete success, I will need to generate visualizations of possible correlations, visualizations of trends, and produce summary tables of the performed statistic analysis. I will flexibly and dynamically adjust the exact criteria for what statistic calculations best answer the proposed questions, but some statistic result will be calculated to determine what correlations and trends exist.

C. Design of Data Analytics Solution

C.1 Hypothesis

1) Some campuses are beginning to receive fewer applications for the doctoral graduate life sciences programs.

2.) Campuses with higher rent costs in their cities receive fewer program applications.

C.2 and C.2.A Analytical Method

All data will be gathered as CSVs and initially have some basic cleaning performed in Excel. These CSVs will then be loaded into dataframes using Python and assessed for further cleaning. Descriptive analytic techniques will be utilized to perform investigation of the data and determine correlations and trends through the inspection of p-values and correlation coefficients. Specifically, I will be performing Man-Kendall tests and Pearson Correlation tests to generate correlations and p-values. These are appropriate methods for the determination of trends and correlations like the ones I am investigating and these methods are readily deployable for a dataset such as this. The matplot, Scipy, Pandas, and Numpy packages will all be utilized for this analysis.

C.3 Tools and Environments

I will be using Microsoft Excel to gather data and perform some data cleaning. I will also utilize Jupyter notebooks along with Python and its many analytic libraries to gather data from CSVs into a dataframe, and to perform data wrangling and cleaning. Python and its plethora of packages will be used to analyze data and visualize outcomes.

C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance

Since I will be investigating trends and correlations, the primary metric I will be employing is the p-value. This is a simple, easily deployable, and broadly accepted metric to determine significance from the null hypothesis – that there is no significant correlation between these data sets. If the deployed analytic techniques produce p-values of <0.05, I will consider the correlations and/or trends associated to be significant.

The p-values will be generated from Mann-Kendall tests on time series data, and will represent the significance of Kendall Tau values representing trends in the time series data.

P-values will also be generated to show the significance of correlation coefficients generated by Pearson Correlation tests, testing the correlation between two data sets (rent cost and application numbers).

C.5 Practical Significance

The practical significance of the analysis performed will be a more robust understanding of the role that housing costs play in the success of graduate programs in the UC system, measured by the application characteristics of the life science doctoral graduate programs across campuses. This information will directly apply to decision-making at the administrative level when understanding the financial needs of graduate students.

C.6 Visual Communication

The data visualizations created for this project will include a series of line graphs showing the trends of application characteristics for each campus between 2015 and 2022, tables containing the output of statistic tests, and a bar graph showing the correlation values campuses with significant correlations between rent cost and application rates.

Since there are several campuses being investigated, the trend lines will be figures containing a series of graphs – one for each campus.

All of these visualizations will be generated with Python and its many analysis driven packages.

D. Description of Dataset

D.1 Source of Data

All data is publicly available. I will draw data from the University of California Information Center, which contains public state data sets. (https://www.universityofcalifornia.edu/)

I will also use data from the Zillow Observed Rent Index (ZORI), which is available for free. (https://www.zillow.com/research/data/)

D.2 Appropriateness of Dataset

The data from the UC information center is direct information from the UC about its own graduate programs. The raw data contains information on applications, admits, new enrolees, admit rates, yield rates, and demographic information on applications. It should be noted that since many students apply to multiple campuses, this data set counts applications rather than applicants.

For the housing cost data, I have chosen to use the Zillow Observed Rent Index. Since we are considering the housing cost of students, I believe this is much more applicable than property values or real estate data – since it is reasonable to believe that the vast majority of students are renting rather than buying houses. This index also adjusts for the availability of rental units and accounts for repeat listings that would otherwise skew the data.

D.3 Data Collection Methods

Data from the University of California was dowladed as CSVs from https://www.universityofcalifornia.edu/about-us/information-center/graduate-admissions.

Data for the ZORI was downloaded in CSV format from https://www.zillow.com/research/data/.

D.4 Observations on Quality and Completeness of Data

The UC data is of high quality, but is only complete back to 2015. There is also some data that I will have to remove during the first cleaning step in Excel, such as demographic data for students, and data for Master's programs, and data for programs other than life sciences.

The ZORI data contains some minor quality issues for my purposes. Foremost, the data is separated into monthly data while I am interested in looking at annual trends in comparison to the annual data from UC. I will have to average this data into annual values before utilizing it. Also this data is only complete to the year 2015, so trends will have to be cut to the last 8 years. There are also missing values for some cities of interest. While most of my cities of interest have few missing values, one city of interest (Riverside, CA) has almost no values at all.

D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances

Because all data is publicly available either through the state of California or through Zillow's research website, it is free to use under the Creative Commons Attribution Non-Commercial No Derivatives 4.0 International Liscense. None of the used data includes any personally identifiable information and contains no privacy or security concerns. There are no inherent ethical considerations concerning this data, and no special precautions are necessary since this data is anonymized and no human interaction was involved in its collection from the publicly available sources.

References

Joo, S.-H., Durband, D. B., & Grable, J. (2008). The Academic Impact of Financial Stress on College Students. Journal of College Student Retention: Research, Theory & Practice, 10(3), 287-305. https://doi.org/10.2190/CS.10.3.c

Clark, J. (2022, September 19). Methodology: Zillow observed rent index (Zori) - Zillow Research. Methodology: Zillow Observed Rent Index (ZORI). https://www.zillow.com/research/methodology-zori-repeat-rent-27092/

Webber, K. L., & Burns, R. (2020). Increases in graduate student debt in the US: 2000 to 2016. Research in Higher Education, 62(5), 709–732. https://doi.org/10.1007/s11162-020-09611-x