**Grants Division (GD) Application Analysis: Does Our Grant Selected Process Potentially have Implicit Bias by Region or State?**

**Introduction**

Grants Division has an investment portfolio of over $1.2 Billion USD supporting a diverse array of agriculture sectors through over 13 grant programs. GD has an application review process of external panels reviews which then rate each application with a score of 1 to 100. GD then takes those scored evaluations and uses them for our own internal review which may include priority selected of application based on project focus, amount of funding available, and our own qualitative review of potential selected applicants for funding. The question GD wants to ask ourselves is: Could there possibly be implicit bias within our application process, specifically by region or state.

I will look at our most competitive program’s (Programs: LFPP, FMPP, and RFSP) grant applications from 2021 to 2023 and provide some descriptive statistical analysis and visualizations on applications and projects awarded by region, state, program, and entity type; I will give deeper and look at average applicant scores/acceptance rates by state for each year, and finally will do some even deeper regression analysis and find out if there is significance difference in applicant scores/acceptance rates by state and region. This will help us answer our question of bias.

**Data Flow and Methodology**

First let’s talk about the data flow: All our grant’s processes are through an external front and backend grants management systems called Grant Solutions. This includes the application submissions, review, and selection. Data is collected on the backend of this external platform via Amazon RedShift. GD has a VPN tunnel connecting this backend data to our own EDAPT Hive Datawarehouse. From the EDAPT Impala ODBC Driver we can pull application data from the Grant Solutions schema to just about any software of system internally.

The Data Flow Steps Are:

1. I will connect the EDAPT Application table from the Schema to Python.
2. Once in Python, I will clean and transform that data to provide two different outputs. One output is the complete dataset and the second is a table breaking down average acceptance rates by state per fiscal year.
3. Create a Scheduler to automate the Python data process. This will allow the python script to be opened and ran on a schedule. The schedule will export data from Python to excel tables.
4. I will then setup live connection from Tableau to the export outputs from Python and conduct further analysis and dashboarding.

The methodology of analysis is to look at statistical summary level information, examine scatter plots and standardized distributions, Examine K-means clusters, and then finished with an OLS regression where the dependent variable is acceptance rates and the independent variables include number of applications, application review score, and states/regions.

**Final Deliverables**

1. Code Snippets from Python as well as output of analysis.
2. Tableau Dashboard tool to help users explore their own analysis of the thesis question.

**Outcome summary**

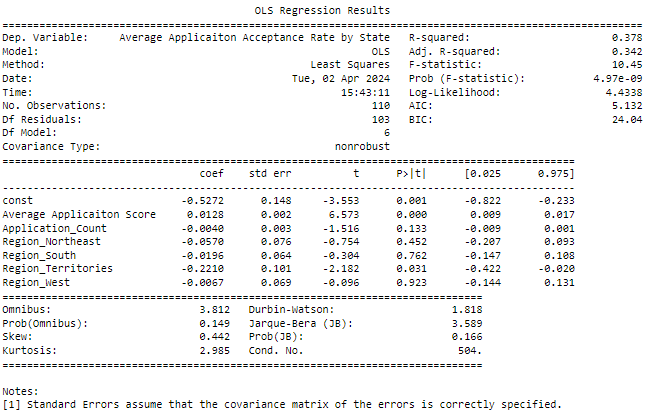
While exploring the data through stats summary analysis, scatter plotting, standardized distributions, K-means analysis, and regression analysis using python and Tableau, it appears that there are no major statistically significant differences in acceptance rates between regions and most States. The US. Territories were the only significant difference in region, there the representation within the data for those territories is too small to make any inference using the data given. As for States, there are some states in which have a p-value less the .005, but they are few and far between and the R-squared for that regression is .664. Though within this analysis I do not feel confident in answering “Does Our Grant Selected Process Potentially have Implicit Bias by Region or State?” as we only looked at three major programs and further analysis is needed. With that being said, the data we do have does provide a glimpse into the possible answer that we are not biased by region or state but is just the beginning.

**What next?**

While exploring the data I saw the lack of observations within average acceptance rates for states by fiscal year. To get more representation/observations, I would incorporate more programs than just the original three and get averages for states by program and fiscal year. This would make for a more robust analysis output and better help see any correlations. I would also dig deeper into more complex multilinear regression analysis and time series analysis.

**Output Analysis:**

Region Regression Output:



Graphical user interface, application, Excel

Description automatically generated**Tableau Application Dashboards**

Chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence