Analysis of Federal Grants for Minority Business Enterprises for the Office of Congressman Joe Kennedy III - Data Report

Yong Zhou, Yue Yu and Ghislaine de Bree

Advisor: Professor Lance Galletti
Client (on behalf of the Congressman's Office): John Merfeld

Introduction / Proposal Suggestions

As per the project description, we are looking into the number of federal grants awarded from top granting agencies for over 10 years and to whom they are awarded. The description listed these agencies as the sources of the grants into which we would look.
  - Massachusetts Department of Transportation (DOT)
  - Small Business Innovation Research (SBIR)
  - US Department of Housing and Urban Development (HUD)
  - US Department of Health and Human Services (HHS)
  - Federal Emergency Management Agency (FEMA)
  - Environmental Protection Agency (EPA)
  - National Science Foundation (NSF)

Then, the crux of the project is to look into the minority owned businesses that are being awarded these grants. There were a few potential data sources that we could use to identify which businesses were owned by minorities (and in particular Black and LatinX owned businesses). The two main suggestions are the Small Business Association (SBA) and the US Black Chamber of Commerce (BCC), both of which collect data on minority owned businesses across the states. The SBA aims to support small business across the country and the BCC does the same thing, but specifically for African-American owned businesses. Given that these were are primary sources for identifying which businesses are minority owned, we were disappointed to find that both of these associations do not release their data publically.

The other option for identifying minority owned businesses, as suggested by the proposal, was to use data from Massachusetts' Supplier Diversity Office's (SDO) Directory Certified Businesses data. This directory has different certifications for different types of MBEs (such as Women, Disadvantaged, Veteran and LGTB Owned Businesses) and we decided that the best way forward was to cast a wide net and download data for businesses that are certified as minority owned by the SDO. After doing so we had data on minority owned business, but no indication of whether or not a specific business is Black or Latinx owned. As we felt this to be the main question that the proposal asked, we reached out to John (who is acting on behalf of the client for our project) with this issue and he suggested that we implement the well used Ethnicolr package as a way to predict the races of these minority owned businesses.

Data Retrieval

In order to get our primary data we used the USA Spending's API to retrieve all federal grants awards in the fiscal years 2010 to 2019 (1st October 2009 to 30th September 2019). We also had to filter our search to include the specified agencies, as listed above. This was not so simple as we found that some of the granters are sub-agencies of other greater federal agencies or not even in the dataset because they were state not federal agencies. We easily got results from the EPA, HUD, HHS and the NSF because these are all federal agencies, and are all defined as top tier agencies in the dataset. The SBIR is a

sub-agency of the Small Business Association (SBA) so we had to scrape for all the data for the SBA and remove non-relevant sub-agencies. The same thing had to be done for FEMA, which is a sub-agency of the Department of Homeland Security. We were unable find data for MassDOT on the USA spending website because this is a state level agency, not a federal. As we felt it important to look into how grants are awarded in the transportation sector, we chose to use grants awarded by the Federal Department of Transportation to businesses owned in Massachusetts.

The request to create a dataset to the API returned four files, that is all assistance prime awards, all assistance subawards, all contracts prime awards and all contracts subawards. Each file includes the different information for different companies and awards. Initially, we thought to only use assistance awards, and not contract awards, but were greatly struggling to find race data for the recipients, as there was no race indication in these files. This is where we would have benefitted from being able to combine our cleaned SDO MBE dataset (with the race predictions from the Ethnicolr package). However, after trying to merge the data on the DUNs number (a code given to businesses that receive federal funds) we found that only 5 of the certified minority owned businesses were present in the federal funds dataset. We also tried to merge on the businesses' names and the first line of their addresses but found no more matches. Clearly, we were missing a lot of data using the SDO certification data (we thought this may be because this certification has to be applied for) and we could not carry out analysis on the assistance awards. After consulting with John, we decided to only look into contracts, as this data has a range of race and minority indicators. The project proposal does seem to lean more towards contracts than assistance. We briefly looked into the contract subawards data, but found that there were no race indicator attributes so we decided to only look at contracts prime awards.

Data Cleaning
In the processes of finding external race data for the assistance awards, we cleaned the SDO MBE data, but ultimately did not find it useful so we have decided to exclude it from this report. Therefore only our main dataset, all prime contracts, was cleaned. The proposal had a list of attributes to include and we found appropriate matches in the data. These included Company Name, Grant Title, Agency, Amount, Date, Recipient Type and Recipient industry. Fortunately, this data had a really extensive list of attributes (with 263) so we decided to include multiple potential sources for some so that we could better look into the conclusions that we could draw from them. We used the awards' unique key as the identifier for each grant, and Company Name and Award Amount were self explanatory. For each grant, the funding agency and funding sub agency were included, so we decided to create a new attribute out of these for the relevant agency listed above (so that we could have all of our agencies of interest in the same attribute). For Date, we decided to include both start and expected end date, so we could potentially look into the length of contracts given to minority owned businesses, and how that developed over the 10 year period. We could not find an attribute for Grant Title, so we used an award description attribute (though these are fairly un-standardized). We used NAICs codes (a system to classify businesses by industry) as markers for the recipients industries, including both the code and the corresponding description. There was a group of indicator attributes for the recipient types (like non-profit, municipal, small business, corporation, institute for higher learning) but we did not want to change this into a single attribute with descriptions because there may be overlap between the many descriptors (there are more than 50 specific attributes). We decided to keep all of these to start with because we felt that we could not arbitrarily chose which attributes would be useful, without looking at their interactions with the race variables first.