The Green Card: A Pathway To Living And Working In The United States

August 2022

Table of Contents

Abstract

Research Methodology

Data Sourcing

Data Cleansing and Preprocessing

Statistical Insights

Employer Related Statistics

Application and Job-Related Statistics

Employee Related Statistics

Machine Learning Models

Continuous Variable Models

Categorical Variable Models

Conclusion

Abstract

This report analyzes trends and relationships between features of prospective employers, employees, and jobs to determine how a hopeful immigrant can have the most successful chance in obtaining a Green Card through an employer. Every year, millions of immigrants become lawful permanent residents of the United States, also known as Green Card holders. In the FY 2020, over 60% of all successful Green Card applicants were either an immediate relative of U.S. citizens or related to a U.S. Citizen or Green Card holder. The most common method, outside of the family-based immigration method, is through the employment-based immigration method. Approximately 21% of new Green Card holders were sponsored by an employer or self-petitioned through investing to create U.S. jobs.

The study compiles, cleans, and manipulates employment PERM (Green Card) data from the U.S. Department of Labor over FY 2015-FY 2021. The report details the natural language preprocessing, feature engineering, cyclical features encoding, normalization techniques, anomaly and outlier detection, statistical analysis and machine learning models used to learn more about permanent immigration through employment.

An applicant should seek out jobs that have a minimum education level of a bachelor’s degree, as jobs that require a bachelor’s degree are 22% more successful than jobs that require a High School diploma or an associate degree. Software developer – applications, software developer – systems software, and computer systems analyst have an 8% advantage over other jobs. Applicants to a large company, with 10,000 or more employees, are 9% more likely to obtain a Green Card than an applicant to a small company, less than 1,000 employees. Unfortunately, jobs that require formal training that are not for a hospital or university have a 62.29% success rate, compared to 94% for jobs at a hospital or university.

The objective of the report is to provide an independent study into how an immigrant can have the most successful chance at becoming permanently employed in the United States. This is done through analyzing and interpreting the trends found within the total PERM data sets. It is important to note, however, the changing nature of immigration and employment; what was the trend over the last seven years, may not repeat itself.

Research Methodology

Data Sourcing

This study examines data from FY2015 – FY2021, sourced from the U.S. Department of Labor PERM Performance Data. The data can be found at this link under Historical Case Disclosure Data: <https://www.dol.gov/agencies/eta/foreign-labor/performance>

The U.S. Department of Labor provides every application from each FY in a separate Excel file. I manually compiled all seven files along the following 69 columns, some of which were derivatives of existing columns:

|  |  |
| --- | --- |
| CASE\_NUMBER | FIRST\_NEWSPAPER\_NAME |
| DECISION\_DATE | FIRST\_AD\_TO\_APP\_DAYS |
| CASE\_STATUS | SECOND\_NEWSPAPER\_AD\_NAME |
| RECEIVED\_DATE | SECOND\_ADVERTISEMENT |
| DAYS\_TO\_DECIDE | SECOND\_AD\_TO\_APP\_DAYS |
| REFILE | JOB\_FAIR |
| EMPLOYER\_NAME | ON\_CAMPUS\_RECRUITING |
| EMPLOYER\_CITY | EMPLOYER\_WEBSITE |
| EMPLOYER\_STATE\_PROVINCE | PRO\_ORG\_AD |
| EMPLOYER\_COUNTRY | JOB\_SEARCH\_WEBSITE |
| EMPLOYER\_NUM\_EMPLOYEES | JOB\_SEARCH\_WEBSITE\_DUR |
| EMPLOYER\_YEAR\_COMMENCED\_BUSINESS | PVT\_EMPLOYMENT\_FIRM |
| PW\_SOC\_CODE | EMPLOYEE\_REF\_PROG |
| PW\_SKILL\_LEVEL | CAMPUS\_PLACEMENT |
| PW\_WAGE | LOCAL\_ETHNIC\_PAPER |
| PW\_UNIT\_OF\_PAY | RADIO\_TV\_AD |
| PW\_DURATION | EMP\_RECEIVED\_PAYMENT |
| APP\_TO\_PW\_DETERMINE | BARGAINING\_REP\_NOTIFIED |
| WAGE\_OFFER\_FROM | POSTED\_NOTICE\_AT\_WORKSITE |
| WAGE\_OFFER\_UNIT\_OF\_PAY | LAYOFF\_IN\_PAST\_SIX\_MONTHS |
| WORKSITE\_CITY | US\_WORKERS\_CONSIDERED |
| WORKSITE\_STATE | COUNTRY\_OF\_CITIZENSHIP |
| JOB\_TITLE | FOREIGN\_WORKER\_BIRTH\_COUNTRY |
| MINIMUM\_EDUCATION | CLASS\_OF\_ADMISSION |
| MAJOR\_FIELD\_OF\_STUDY | FOREIGN\_WORKER\_EDUCATION |
| REQUIRED\_TRAINING | FOREIGN\_WORKER\_INFO\_MAJOR |
| REQUIRED\_EXPERIENCE | FOREIGN\_WORKER\_YRS\_ED\_COMP |
| ACCEPT\_FOREIGN\_EDUCATION | FOREIGN\_WORKER\_INST\_OF\_ED |
| ACCEPT\_ALT\_OCCUPATION | FOREIGN\_WORKER\_TRAINING\_COMP |
| JOB\_OPP\_REQUIREMENTS\_NORMAL | FOREIGN\_WORKER\_REQ\_EXPERIENCE |
| FOREIGN\_LANGUAGE\_REQUIRED | FOREIGN\_WORKER\_ALT\_ED\_EXP |
| PROFESSIONAL\_OCCUPATION | FOREIGN\_WORKER\_ALT\_OCC\_EXP |
| APP\_FOR\_COLLEGE\_U\_TEACHER | EMPLOYER\_COMPLETED\_APPLICATION |
| SWA\_DURATION | NAICS\_CODE |
| SUNDAY\_EDITION\_NEWSPAPER |  |

DAYS\_TO\_DECIDE, SWA\_DURATION, and JOB\_SEARCH\_WEBSITE\_DUR were derived from start to end dates. The FIRST\_AD\_TO\_APP\_DAYS and SECOND\_AD\_TO\_APP\_DAYS were derived from the start date and the application received date.

Other columns, mostly related to advertisement source, may have been turned into Y or N columns based on if a Date value existed in the column, e.g., JOB\_SEARCH\_WEBSITE was engineered from the start date, if a start date existed then the date value turned into a Y. After the initial compiling of data, there were 737,759 applications and 69 columns.

Moving forward, this paper will be written with the assumption that the reader has looked at the terminology and layout of the PERM Green Card application, which can be found on the same webpage the data is stored in.

Data Cleansing and Preprocessing

As with anything of human nature, there were erroneous answers to some of the questions found on the Green Card PERM form. To have the most accurate insights, features, and outputs, the data was cleaned extensively.

PW SOC CODE and Feature Engineering the Case Status Outcome

The first feature cleaned was the PW SOC Code. Any applicants that did not have a “-“ in the code was immediately dropped as it did not follow the PW format and any inferences made to complete the code would not be accurate. 1,104 applicants were dropped for not having a “-“.

Some PW codes had decimals at the end of the code. Decimals were an older standard that are not used anymore. The U.S. Department of Labor moved to consolidate PW Codes by removing the decimal, and so the 2,844 applicants with decimals in their PW Code had their decimals removed.

Due to the manual compiling of seven years of data, some of the PW codes turned into dates. This was cleaned by splitting the Code at the “-“. Any “Nov” to the left of the “-“ were replace with 11. If a Code had “Nov” to the right of the “-“ it would have two digits to the left of the “-“. The “Nov” on the right of the “-“ were replaced with the two digits to the left of the “-“ and the two digits to the left of the “-“ were replaced with 11. After both parts of the split Codes with “Nov” were cleaned, two zeros were added to the end to match the PW SOC Code format.

“Nov” to the left of the “-“: Nov-12 > 11-12 > 11-1200

“Nov” to the right of the “-“: 12-Nov > 11-12 > 11-1200

Following the PW SOC Code, the Case Status column was turned into dummy variables after light feature engineering. Certified Expired and Certified were a success and defined as a 1, Withdrawn or Denied were a rejection and defined as a 0.

Wage

Next, the wage and unit of wage were cleaned. Any rows without a Wage Offer, Unit of Wage Offer, PW Wage Offer, or Unit of PW Wage Offer were dropped. This led to 1,236 rows dropped and an additional 6 rows dropped that had “########” as the Wage Offer and the PW Wage Offer.

The Wage Offer Unit and the PW Wage Offer Unit were both abbreviated such that YEAR became YR, MONTH became MTH, HOUR became HR, WEEK became WK, and BI-WEEKLY became BI.

To compare and predict annual wage offered, both Wage Offer and PW Wage Offer had to be annualized. This was more laborious than expected as many applications were not completed accurately. E.g., an applicant would have a Wage Offer of $50,000 but the Unit was WK, when it was supposed to be YR.

There were 1,460 instances where the annualized Wage Offer was less than the annualized PW Wage offer, which is not possible as the U.S. Government requires all Green Card Sponsored Applicants get paid more than the PW Wage.

Statistical Insights

Employer Related Statistics

Application and Job-Related Statistics

Employee Related Statistics

Machine Learning Models

Continuous Variable Models

Categorical Variable Models

Conclusion