USA Visa Analysis

Using Apache Spark

Prepared by:

Akanksha Utreja 11910056

Devashish Dhiman 11910075

Swapnil Vermani 11910052

3 June 2018

Big Data Management by *Prof. Peeyush Taori*

INDEX

ABSTRACT

A permanent labor certification issued by the Department of Labor (DOL) allows an employer to hire a foreign worker to work permanently in the United States. In most instances, before the U.S. employer can submit an immigration petition to the Department of Homeland Security's U.S. Citizenship and Immigration Services (USCIS), the employer must obtain a certified labor certification application from the DOL's Employment and Training Administration (ETA). The DOL must certify to the USCIS that there are not sufficient U.S. workers able, willing, qualified and available to accept the job opportunity in the area of intended employment and that employment of the foreign worker will not adversely affect the wages and working conditions of similarly employed U.S. workers.

Data covers 2012-2017 and includes information on employer, position, wage offered, job posting history, employee education and past visa history, associated lawyers, and final decision.

ABOUT THE DATASET

Data covers 2012-2017 and includes information on employer, position, wage offered, job posting history, employee education and past visa history, associated lawyers, and final decision.

The CSV file contains data of 42,194 applicants across . The data is rich and spread across the occupation of each applicant, their position, education, final decision and more. Along with that the data is historically rich with the job hosting history and applicant's past visa history. This 330 MB data is really tough to be accessed on Excel let alone perform computational and hence we would be using Spark SQL for the computation

The few primary column of Data relevant for this EDA is mentioned below.
1. Wage Offered
2. Application Case status
3. Country of Citizenship
4. Applicant Name
5. Job training information

PROJECT INTRODUCTION

The primary reason of picking this project was to observe any trend with VISA application process of USA Government's Department of Labor. This data was rich with indexes over the applicant information. It was a good dataset as we could inference a lot from the profile of applicant. Although we faced few challenges when we started to analyse this data. The primary being the sheer size of data. It is 330 MB size of data with 42,194 Rows and 154 Columns. It became a tough task for Excel to handle, and as per our intention of mining the data using SQL operation would be a little too much for Excel. Hence we decided to Using Spark SQL for this operation. Spark is a Big data Technology, built on the principles of Hadoop.

Having said that, Spark also had some limitations we had to overcome. As you would know, in a traditional set-up of Big Data project. We store the data in an HDFS system in form of RDDs and then use spark technologies we compute and analyse this data, however, 330 MB was using a lot of local computational memory. Hence we decided with going with Cloud based computation. We picked Databricks, a big data tool based on Python and built by Apache.

It offers us a DBFS, which is data bricks file system, which is a cloud based file storage system. We can use the file stored on their server to compute thus reducing the location computation power requirement. This helped us perform faster computation and analysis of the file.

We would also use Tableau for the visualisation. Spark is really good with Computation but is not the best tool when it comes to the Visualisation. Here it begins.

BUSINESS QUESTIONS IDENTIFIED

The following questions were identified by the team after multiple iterations and observations

1. Acceptance on the basis of wages

2. Acceptance on the basis of Country of Origin

3. Acceptance on the basis of Company Reputation

4. Acceptance and Rejection over the years

5. Acceptance on the basis of Work Experience

6. Profile of the applicants

7. City with the maximum number of visa applications

8. Industry with the maximum number of visa applications

APPLICATION ACCEPTANCE ON THE BASIS OF WAGES

We would be analysis the impact of wages on the acceptance of rejection of the visa application. With this We would be able to identify another metrics, the maximum or minimum wages for accepted Applicants.

Approach:

We have calculated the max and min wage for the applications where the case_status is "CERTIFIED".We had to convert it to INT as the wages were String.

Maximum wages:

```
max_wage=sqlContext.sql("Select MAX(INT(wage_offer_from_9089)) from usvisa where case_status='Certified'")
max_wage.show()

+-----------------------------------+
|max(CAST(wage_offer_from_9089 AS INT))|
+-----------------------------------+
|                           9582600|
+-----------------------------------+
```

Minimum wages:

```
min_wage=sqlContext.sql("Select MIN(INT(wage_offer_from_9089)) from usvisa where case_status='Certified'")
min_wage.show()

+-----------------------------------+
|min(CAST(wage_offer_from_9089 AS INT))|
+-----------------------------------+
|                                 6|
+-----------------------------------+
```

Insights:

We can see here applications with wages as low as 6 has made the cut of being accepted.On the other hand the upper limit sits at 9582600. So , it is safe to say that visa acceptance caters to a varied range of wages and maybe other factors influences acceptance a bit more.

APPLICATION ACCEPTANCE ON THE BASIS OF COUNTRY OF ORIGIN

Our objective here is to identify the correlation between birth country and results.

Approach:

We have considered birth country as country_of_citizenship (the best suited metric).Then we have calculated the no of applications from every country , the acceptances from that country and the rejections from the country.We have further decided some metrics called the acceptance_ratio which is the total number of acceptances/total number of applications and the rejection_ratio which is the total number of rejections/total number of applications.

Acceptance Count

```
acceptance=sqlContext.sql("Select Count(case_status) as acceptance_count ,country_of_citzenship from usvisa where case_status='Certified' group by
country_of_citzenship order by acceptance_count DESC")
acceptance.show(10)
+----------------+---------------------+
|acceptance_count|country_of_citzenship|
+----------------+---------------------+
|          170126|                 null|
|            6226|                INDIA|
|             652|                CHINA|
|             645|               CANADA|
|             620|          SOUTH KOREA|
|             367|          PHILIPPINES|
|             317|               MEXICO|
|             184|       UNITED KINGDOM|
|             164|             PAKISTAN|
|             154|               TAIWAN|
+----------------+---------------------+
only showing top 10 rows
```
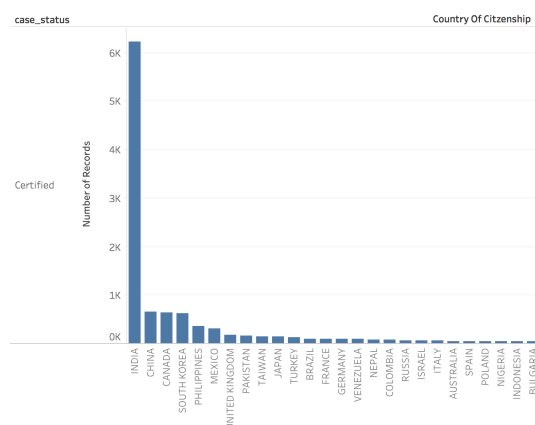

Rejection Count

```
rejection=sqlContext.sql("Select Count(case_status) as rejection_count ,country_of_citzenship from usvisa where case_status='Denied' group by country_of_citzenship
order by rejection_count DESC")
rejection.show(10)
+---------------+---------------------+
|rejection_count|country_of_citzenship|
+---------------+---------------------+
|          22966|                 null|
|            972|                INDIA|
|            245|               MEXICO|
|            214|          PHILIPPINES|
|            203|          SOUTH KOREA|
|            113|                CHINA|
|             85|               CANADA|
|             48|              ECUADOR|
|             48|             PAKISTAN|
|             47|       UNITED KINGDOM|
+---------------+---------------------+
only showing top 10 rows
```
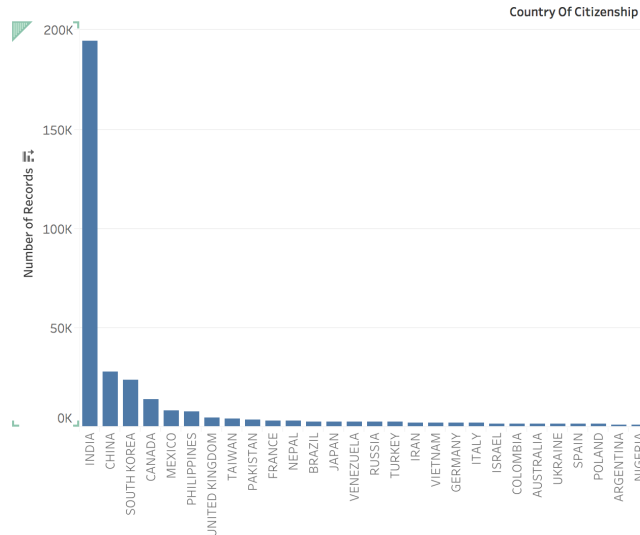
Application Count

```
application=sqlContext.sql("Select Count(case_status) as application_count ,country_of_citzenship from usvisa  group by country_of_citzenship order by
application_count DESC")
application.show(10)
```

```
+-----------------+--------------------+
|application_count|country_of_citzenship|
+-----------------+--------------------+
|           353788|                null|
|            10547|               INDIA|
|             1147|         SOUTH KOREA|
|             1106|               CHINA|
|             1028|              CANADA|
|              748|         PHILIPPINES|
|              677|              MEXICO|
|              322|      UNITED KINGDOM|
|              282|            PAKISTAN|
|              272|              TAIWAN|
+-----------------+--------------------+
only showing top 10 rows
```


Country Of Citizenship

Acceptance Ration for countries having more than 100 applications

```
import pyspark.sql.functions as F

acceptance_ratio = application\
    .join(acceptance, "country_of_citzenship")\
    .withColumn("ratio", (F.col("acceptance_count") / F.col("application_count")))


acceptance_ratio.sort('ratio', ascending=False).show()
acceptance_ratio.sort('ratio', ascending=False).filter(F.col("application_count") > 100).show(10)
```

```
+--------------------+-----------------+----------------+------------------+
|country_of_citzenship|application_count|acceptance_count|             ratio|
+--------------------+-----------------+----------------+------------------+
|              CANADA|             1028|             645| 0.627431906614786|
|               ITALY|              101|              63|0.6237623762376238|
|              TURKEY|              223|             137|0.6143497757847534|
|               JAPAN|              235|             140|0.5957446808510638|
|               INDIA|            10547|            6226|0.5903100407698871|
|               CHINA|             1106|             652|0.5895117540687161|
|             GERMANY|              162|              95|0.5864197530864198|
|              ISRAEL|              113|              66| 0.584070796460177|
|            PAKISTAN|              282|             164|0.5815602836879432|
|              BRAZIL|              175|             101|0.5771428571428572|
+--------------------+-----------------+----------------+------------------+
only showing top 10 rows
```

Rejection Ratio for countries having more than 100 applications at least

```python
import pyspark.sql.functions as F

rejection_ratio = application\
    .join(rejection, "country_of_citzenship")\
    .withColumn("ratio", (F.col("rejection_count") / F.col("application_count")))

rejection_ratio.sort('ratio', ascending=False).show()
rejection_ratio.sort('ratio', ascending=False).filter(F.col("application_count") > 100).show(10)
```

```
+--------------------+-----------------+---------------+-------------------+
|country_of_citzenship|application_count|rejection_count|              ratio|
+--------------------+-----------------+---------------+-------------------+
|             ECUADOR|              109|             48|0.44036697247706424|
|              MEXICO|              677|            245|  0.361890694239291|
|              POLAND|              107|             35|0.32710280373831774|
|         PHILIPPINES|              748|            214|0.28609625668449196|
|            COLOMBIA|              151|             36|0.23841059602649006|
|               ITALY|              101|             21| 0.2079207920792079|
|         SOUTH KOREA|             1147|            203|0.17698343504795117|
|               JAPAN|              235|             41|0.17446808510638298|
|            PAKISTAN|              282|             48| 0.1702127659574468|
|              BRAZIL|              175|             28|               0.16|
+--------------------+-----------------+---------------+-------------------+
only showing top 10 rows
```

Insights:

1. Here we can see that Although India leads in highest number of applications, rejections and acceptances but we cannot conclude anything from that as we have to believe in more normalised metric which is perhaps acceptance ratio and rejection ratio
2. Some countries have 1.0 as the acceptance and rejection ratio because they have less number of applications to support that claim.Eg: 1 acceptance from 1 application.
3. When we analyse only of countries greater than 100 as number of applications , we see that Canada has the highest acceptance ratio which obvious as Canada being the neighbouring country and Ecuador has the highest rejection ratio
4. As for India the acceptance ratio is under top 5 counties around 60 % while rejection as low as 9 %.

APPLICATION ACCEPTANCE ON THE BASIS OF COMPANY REPUTATION

We would be analysing what companies holds the maximum Visa Acceptance ration.

Approach

We have considered employer_name (the best suited metric).Then we have calculated the no of applications from every employer, the acceptances from that employer and the rejections from the employer.We have further decided some metrics called the acceptance_ratio which is the total number of acceptances/total number of applications

Application count

```
top_20_applications=sqlContext.sql("Select COUNT(case_number) as app_count,employer_name from usvisa group by employer_name order by app_count DESC")
top_20_applications.show(10)
```

```
+---------+--------------------+
|app_count|       employer_name|
+---------+--------------------+
|    11587|COGNIZANT TECHNOL...|
|     4625|MICROSOFT CORPORA...|
|     4127|   INTEL CORPORATION|
|     3771|         GOOGLE INC.|
|     2716|AMAZON CORPORATE LLC|
|     2097| CISCO SYSTEMS, INC.|
|     1795|         INFOSYS LTD.|
|     1730|          APPLE INC.|
|     1552|ORACLE AMERICA, INC.|
|     1268|      FACEBOOK, INC.|
+---------+--------------------+
only showing top 10 rows
```

Acceptance Count

```
top_20_acceptances=sqlContext.sql("Select COUNT(case_number) as acc_count,employer_name from usvisa where case_status='Certified' group by employer_name order by acc_count DESC ")
top_20_acceptances.show(10)
```

```
+---------+--------------------+
|acc_count|       employer_name|
+---------+--------------------+
|     5367|COGNIZANT TECHNOL...|
|     2360|   INTEL CORPORATION|
|     2196|         GOOGLE INC.|
|     2093|MICROSOFT CORPORA...|
|     1519|AMAZON CORPORATE LLC|
|      992|          APPLE INC.|
|      925|ORACLE AMERICA, INC.|
|      794| CISCO SYSTEMS, INC.|
|      776|         INFOSYS LTD.|
|      729|      FACEBOOK, INC.|
+---------+--------------------+
only showing top 10 rows
```

Rejection Count

```
top_20_rejections=sqlContext.sql("Select COUNT(case_number) as rej_count,employer_name from usvisa where case_status='Denied' group by employer_name order by rej_count DESC ")
top_20_rejections.show(10)
```

```
+---------+--------------------+
|rej_count|       employer_name|
+---------+--------------------+
|      316|   INTEL CORPORATION|
|      279|INFOSYS TECHNOLOG...|
|      256|HOUSE OF RAEFORD ...|
|      254|HOUSE OF RAEFORD ...|
|      122|    HARRISON POULTRY|
|      120|AMERICA'S CATCH, ...|
|       96|NORMAN W. FRIES, ...|
|       85|COGNIZANT TECHNOL...|
|       75|NEW YORK CITY DEP...|
|       75|KOCH FOODS OF CIN...|
+---------+--------------------+
only showing top 10 rows
```
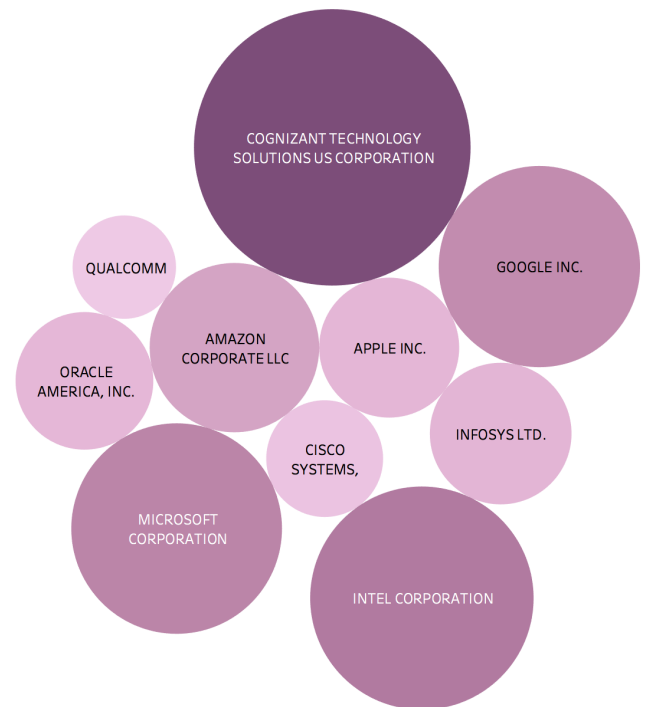
Acceptance Ratio

```
Company_acceptance_ratio = top_20_applications\
    .join(top_20_acceptances, "employer_name")\
    .withColumn("ratio", (F.col("acc_count") / F.col("app_count")))

Company_acceptance_ratio.sort('ratio', ascending=False).show()
Company_acceptance_ratio.sort('ratio', ascending=False).filter(F.col("app_count") > 100).show(10)
```

```
+--------------------+---------+---------+------------------+
|       employer_name|app_count|acc_count|             ratio|
+--------------------+---------+---------+------------------+
|PURE BEAUTY FARMS...|      116|      113|0.9741379310344828|
|        HER SERVICES|      105|      101|0.9619047619047619|
|PATNI AMERICAS, INC.|      129|      123|0.9534883720930233|
|CLEANERS OF AMERI...|      121|      113|0.9338842975206612|
|HEWLETT-PACKARD E...|      110|      102|0.9272727272727272|
|JCG FOODS OF GEOR...|      155|      137|0.8838709677419355|
|MUY HAMBURGER PAR...|      426|      373|0.8755868544600939|
|PricewaterhouseCo...|      225|      197|0.8755555555555555|
|CAPGEMINI AMERICA...|      165|      144|0.8727272727272727|
|ADOBE SYSTEMS INC...|      261|      197|0.7547892720306514|
+--------------------+---------+---------+------------------+
only showing top 10 rows
```


Inferences

1. Here we can see that Cognizant has the highest acceptances but its number of applications is also the highest.
2. Hence we get more clearer insight when we compare acceptance ratios and we can clearly see big names with high acceptance ratio like Pure Beauty Farms,HER Services,Patni Americas, INC. PWC, CAPGEMINI.

ACCEPTANCE AND REJECTION OVER THE YEARS

We want to observe whether the acceptance or Rejection rate has been increased or Decreased over the year.

```
date= sqlContext.sql("select distinct(YEAR ( decision_date ) )from usvisa")

rate= sqlContext.sql("Select YEAR( decision_date) as Year, count(case_status) as Acceptance_Count  from usvisa where case_status='Certified' group by Year order by
Acceptance_Count").show(10)

rate= sqlContext.sql("Select  YEAR( decision_date) as Year, count(case_status) as Rejection_Count from usvisa where case_status='Denied' group by Year order by
Rejection_Count").show(10)
```

Acceptance Count

```
+----+----------------+
|Year|Acceptance_Count|
+----+----------------+
|2011|            4747|
|2012|            7060|
|2013|           13687|
|2014|           35470|
|2015|           40018|
|2016|           80951|
+----+----------------+
```

How did the Acceptance Rate change over Years?

Decision Date
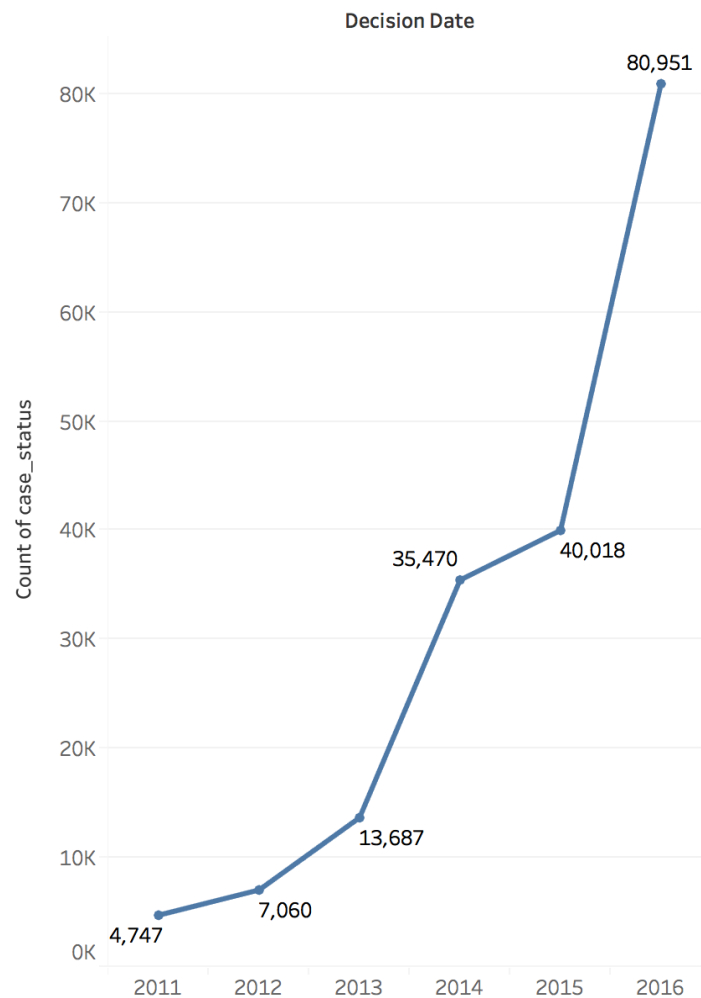
Year and Rejection

Rejection Count

```
+----+---------------+
|Year|Rejection_Count|
+----+---------------+
|2011|           1128|
|2012|           2838|
|2014|           4423|
|2013|           5433|
|2016|           5893|
|2015|           5934|
+----+---------------+
```

Decision Date
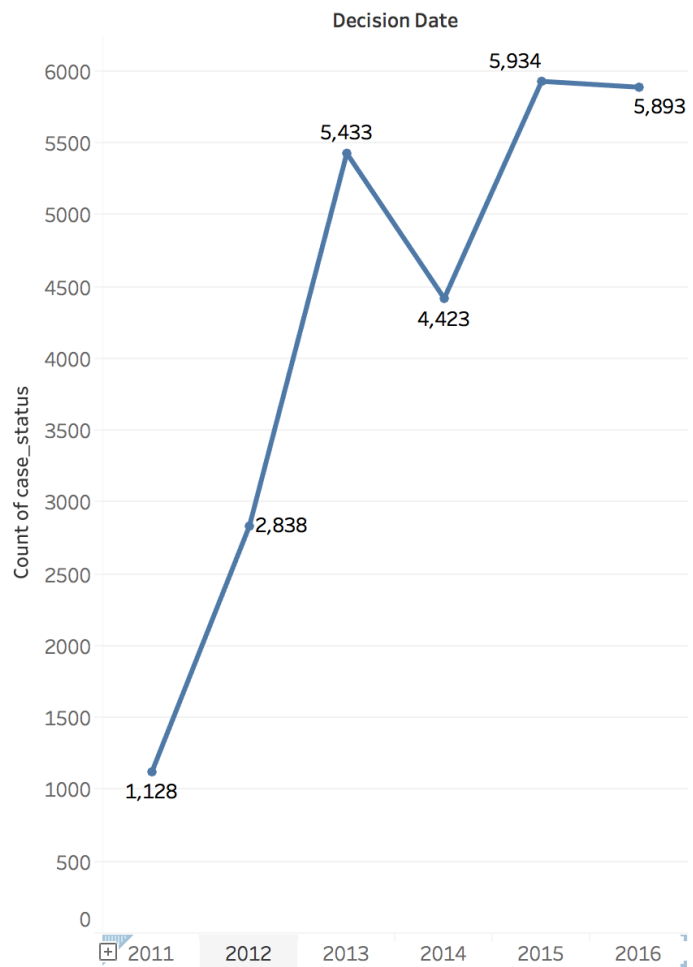
Insights:

1. We can see from the numbers and the visualisations that acceptance ratio have only increased over the years while rejection ratio have been more or less constant.
2. This behaviour could be attributed to the welcoming policies of visa Acceptances over the year by US.

ACCEPTANCE ON THE BASIS OF WORK EXPERIENCE

Approach

- We have taken "job_info_experience_num_months" and "case_status" to find the influence of Work experience on Visa Acceptance.

- First, we have analysed the different time ranges that are present in our data set by writing queries.

- After that, we found how the work experience is affecting the visa acceptance by making a table that shows the count of certified case_status i.e. the visas which were accepted and the work experience.

- To dig deeper, we found the work experience range in which the maximum rejections were happening by creating another table.

```
exp= sqlContext.sql("select distinct(job_info_training_num_months) from usvisa where job_info_training_num_months !=''")
# Impact of work experience on Visa acceptance

WorkExpAcceptance= sqlContext.sql("Select job_info_training_num_months, count(case_status) as Acceptance_Count  from usvisa where case_status='Certified' and
job_info_training_num_months!='' group by job_info_training_num_months order by  Acceptance_Count DESC").show(10)

WorkExpAcceptance= sqlContext.sql("Select job_info_training_num_months, count(case_status) as Rejection_Count  from usvisa where case_status='Denied' and
job_info_training_num_months!='' group by job_info_training_num_months order by  Rejection_Count DESC ").show(10)
```

Acceptance Count

```
+---------------------------+----------------+
|job_info_training_num_months|Acceptance_Count|
+---------------------------+----------------+
|                       36.0|            1066|
|                        0.0|             185|
|                       60.0|             170|
|                       48.0|             142|
|                       72.0|             119|
|                       12.0|             106|
|                       24.0|              90|
|                        6.0|              43|
|                        3.0|              34|
|                        1.0|              23|
+---------------------------+----------------+
only showing top 10 rows
```

How does Work Experience impact Acceptance of Visa?
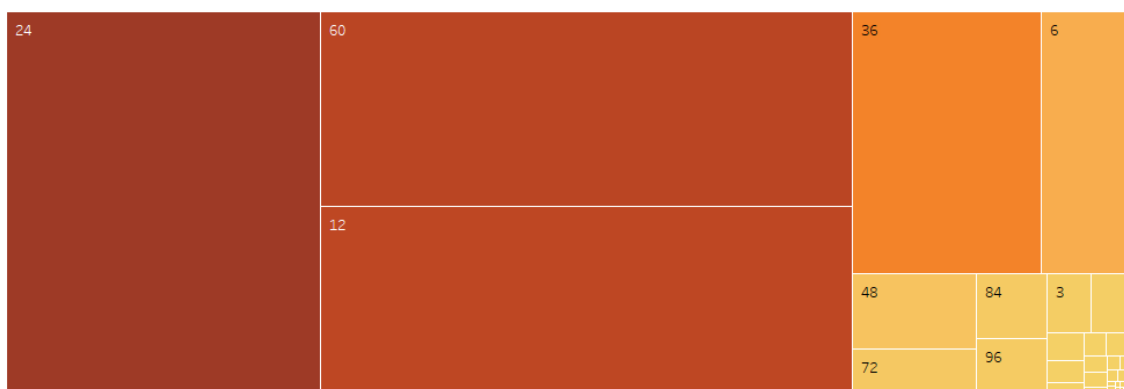
```
+---------------------------+---------------+
|job_info_training_num_months|Rejection_Count|
+---------------------------+---------------+
|                        1.0|            103|
|                       12.0|             80|
|                        6.0|             77|
|                        3.0|             76|
|                       36.0|             71|
|                       24.0|             60|
|                        2.0|             45|
|                       60.0|             23|
|                        0.0|             14|
|                       18.0|              8|
+---------------------------+---------------+
only showing top 10 rows
```

How does Work Experience impact Acceptance of Visa?
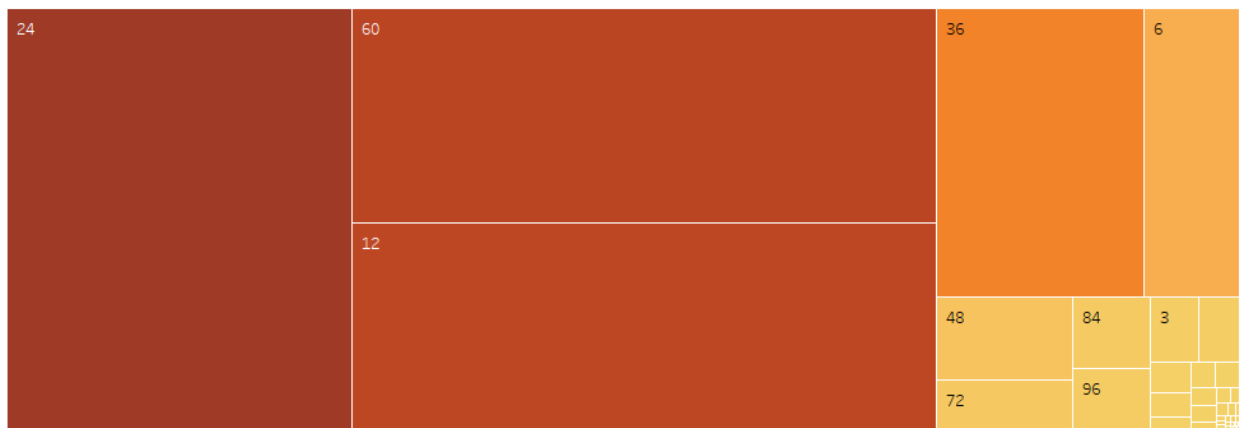
Insights:

1. From the table 1 it is clear that a person having 24 months of experience is in the most ideal situation and has high chances of getting his visa accepted.

2. Then lastly we saw the impact of work experience on Rejections, from the below we can conclude that the person with 24 years of experience is in the danger zone as well. This is because since the maximum applications come from people aged 24 hence, the number of acceptance as well as rejections are maximum for this age group.

PROFILE OF THE APPLICANTS

Approach

- We are taking the columns "pw_soc_title" and "case_status" from the dataset and creating tables for different scenarios to analyse using Spark SQL.

- First, we find who apply the most for US permanent visa's. Is it the doctors, engineers etc.

- After that, we analyse whose applications gets accepted the most.

- Lastly, we find the ratio of the accepted application to the total applications to find the "Acceptance Ratio".

Application count

```
career_application_rate = sqlContext.sql("Select pw_soc_title as career, count(case_status) as app_Count  from usvisa  group by career order by  app_Count DESC")
career_application_rate.show(10)

+--------------------+---------+
|              career|app_Count|
+--------------------+---------+
|Software Develope...|   114841|
|Computer Systems ...|    36700|
|Software Develope...|    21356|
|Electronics Engin...|    13192|
|Computer and Info...|     9631|
|Computer Software...|     6345|
|Accountants and A...|     6122|
|Network and Compu...|     5986|
|Meat, Poultry, an...|     5330|
|Mechanical Engineers|     5260|
+--------------------+---------+
only showing top 10 rows
```

Acceptance count

```
career_acceptance_rate= sqlContext.sql("Select pw_soc_title as career, count(case_status) as acc_count  from usvisa where case_status='Certified' group by career order by  acc_count DESC")
career_acceptance_rate.show(10)

+--------------------+---------+
|              career|acc_count|
+--------------------+---------+
|Software Develope...|    57998|
|Computer Systems ...|    17040|
|Software Develope...|    11329|
|Electronics Engin...|     6591|
|Computer and Info...|     4569|
|Computer Software...|     3013|
|Network and Compu...|     2885|
|Accountants and A...|     2841|
|Mechanical Engineers|     2662|
|Meat, Poultry, an...|     2269|
+--------------------+---------+
only showing top 10 rows
```

Acceptance Ratio

```
#Acceptance ratio
career_acceptance_ratio = career_application_rate\
    .join(career_acceptance_rate, "career")\
    .withColumn("ratio", (F.col("acc_count") / F.col("app_count")))

career_acceptance_ratio.sort('ratio', ascending=False)
career_acceptance_ratio.sort('ratio', ascending=False).filter(F.col("app_count") > 100).show(10)
```

```
+-------------------+---------+---------+------------------+
|             career|app_Count|acc_count|             ratio|
+-------------------+---------+---------+------------------+
|Computer Occupati...|     1554|     1413|0.9092664092664092|
|Information Secur...|      580|      357|0.6155172413793103|
|Human Resources S...|      268|      160|0.5970149253731343|
|Computer and Info...|      125|       74|             0.592|
|   Animal Scientists|      147|       86|0.5850340136054422|
|Architecture Teac...|      103|       60|0.5825242718446602|
|Janitors and Clea...|     1418|      826|0.5825105782792666|
|Farmworkers, Farm...|      112|       65|0.5803571428571429|
|Combined Food Pre...|      998|      579|0.5801603206412825|
|Receptionists and...|      114|       65|0.5701754385964912|
|Transportation, S...|     1065|      603|0.5661971830985916|
|Agricultural Scie...|      101|       57|0.5643564356435643|
|Farmworkers and L...|      484|      272|0.5619834710743802|
|Recreation and Fi...|      128|       71|         0.5546875|
| Compliance Officers|      354|      196|0.5536723163841808|
|        Veterinarians|      338|      187|0.5532544378698225|
|        Statisticians|     3427|     1893|0.5523781733294426|
```

Insights:

1. Who apply the most for US Permanent visa is shown in the table below. It is clear that Software developers apply the most while Computer User Support apply the least.

2. We find the who are the group whose visa's are accepted the most, and from the table we can see that the Software developers are the group of people whose visas are accepted the most.

3. The third table gives us the insights on the Acceptance ration. From the below table, it is clear that the Computer occupation has the best Acceptance Ratio.

CITY WITH THE MAXIMUM NUMBER OF VISA APPLICATIONS

Approach

- We have taken job_info_work_city and case_status, created different tables from these two and drew insights.

- The various tables have been shown in the Findings.

Application Count

```
+-----------------+-----------------+
|job_info_work_city|application_count|
+-----------------+-----------------+
|          New York|            16971|
|   College Station|            11026|
|          San Jose|             8428|
|           Redmond|             8007|
|           Houston|             7354|
|     San Francisco|             7151|
|     Mountain View|             6896|
|       Santa Clara|             5852|
|         Sunnyvale|             5613|
|           Seattle|             5227|
|             Plano|             5157|
|           Chicago|             4503|
|         San Diego|             3941|
|       Los Angeles|             3776|
|            Irving|             3310|
|            Austin|             3162|
|           Atlanta|             3093|
|          Hillsboro|            3063|
```

Acceptance Count

```
+---------------+-----------------+
|acceptance_count|job_info_work_city|
+---------------+-----------------+
|           8971|          New York|
|           5457|   College Station|
|           4028|          San Jose|
|           3867|     San Francisco|
|           3825|     Mountain View|
|           3575|           Houston|
|           3134|           Redmond|
|           2975|       Santa Clara|
|           2872|           Seattle|
|           2714|             Plano|
|           2708|         Sunnyvale|
|           2284|           Chicago|
|           1959|         San Diego|
|           1804|       Los Angeles|
|           1758|            Irving|
|           1735|            Austin|
|           1733|          Hillsboro|
|           1594|          Cupertino|
```

Rejection Count

```
+--------------+-----------------+
|rejection_count|job_info_work_city|
+--------------+-----------------+
|           651|          New York|
|           547|           Fremont|
|           425|           Houston|
|           352|       Los Angeles|
|           276|          NEW YORK|
|           254|     West Columbia|
|           242|         Rose Hill|
|           241|            Dallas|
|           231|          Hillsboro|
|           217|          San Jose|
|           207|         Sunnyvale|
|           205|       Santa Clara|
|           200|           Teaneck|
|           191|           Chicago|
|           187|            Edison|
|           185|          Brooklyn|
|           175|     San Francisco|
|           171|            Irving|
```

```
+------------------+-----------------+----------------+------------------+
|job_info_work_city|application_count|acceptance_count|             ratio|
+------------------+-----------------+----------------+------------------+
|          New York|            16971|            8971|0.5286076247716693|
+------------------+-----------------+----------------+------------------+

+------------------+-----------------+----------------+-------------------+
|job_info_work_city|application_count|rejection_count|              ratio|
+------------------+-----------------+----------------+-------------------+
|          New York|            16971|             651|0.038359554534205406|
+------------------+-----------------+----------------+-------------------+
```

Acceptance Ratio

```
+-------------------+-----------------+----------------+------------------+
|  job_info_work_city|application_count|acceptance_count|             ratio|
+-------------------+-----------------+----------------+------------------+
|Pine Mountain Valley|             141|             124|0.8794326241134752|
|          Lake Mary|              107|              81|0.7570093457943925|
|              PLANO|              169|             121|0.7159763313609467|
|             DOBSON|              224|             158|0.7053571428571429|
|         Lewisville|              186|             129|0.6935483870967742|
|           Anderson|              375|             258|             0.688|
|         PENDERGRASS|             192|             130|0.6770833333333334|
|        Ridge Spring|             142|              96| 0.676056338028169|
|       Moon Township|             404|             269|0.6658415841584159|
|         Mooresville|             102|              67|0.6568627450980392|
|          RICHARDSON|             262|             170| 0.648854961832061|
|              Venice|             119|              76|0.6386554621848739|
|         Auburn Hills|             404|             252|0.6237623762376238|
|          Charleston|             119|              73|0.6134453781512605|
```

Insights:

1. We can say that New York receives the maximum number of applications.
2. We can conclude that New York has the maximum number of Rejections.
3. It is very interesting to note that although New York receives the maximum number of applications and accepts and rejects the maximum number of applications, it does not have the highest acceptance ratio or the rejection ratio.
   1. The maximum Acceptance ratio is for Pine Mountain Valley.
   2. The Acceptance ratio for New York is 0.52
   3. The Rejection ratio for New York is 0.038

## *INDUSTRY WITH THE MAXIMUM NUMBER OF VISA APPLICATIONS*

Approach

- We took two columns naics_2007_us_title as sector and case_status to build different tables and find insights.

- We have created tables for the sectors having maximum acceptances and rejections. It is explained in detail in Findings.
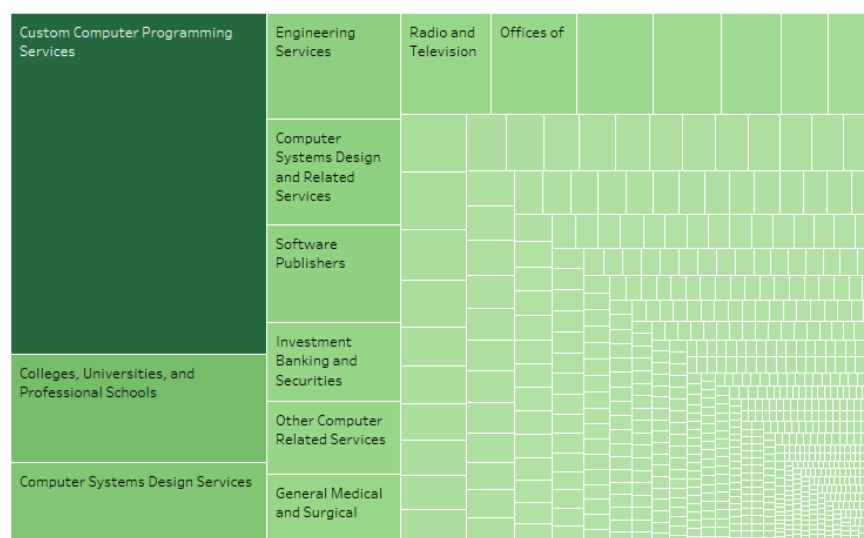
Inferences

- We have created tables for the sectors having maximum acceptances and rejections. It is explained in detail in Findings.

Industry sector recieving max requests
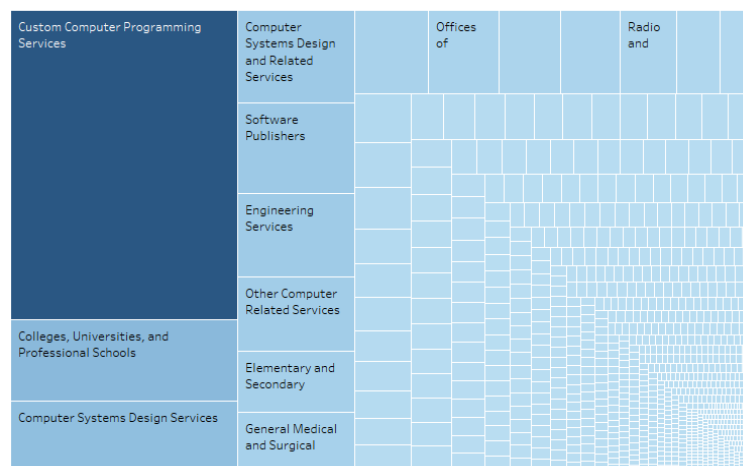
- The below table shows sector having the maximum acceptance count. We can see that it is Custom Computer Programming services.

Max Acceptances

- The below table shows the sector having the maximum acceptance ratio. It is clearly Radio and Television.

```
+-------------------+---------+---------+------------------+
|             sector|app_Count|acc_count|             ratio|
+-------------------+---------+---------+------------------+
|Radio and Televis...|      274|      228|0.8321167883211679|
|All Other Informa...|      111|       83|0.7477477477477478|
|Investment Bankin...|      360|      259|0.7194444444444444|
|  Electronic Shopping|     165|      118|0.7151515151515152|
|Administrative Ma...|      212|      151|0.7122641509433962|
|Colleges, Univers...|     1067|      690|0.6466729147141518|
|Offices of Physic...|      341|      216|0.6334310850439883|
|Offices of Certif...|      149|       94|0.6308724832214765|
|Engineering Services|      573|      354|0.6178010471204188|
|General Medical a...|      397|      235|0.5919395465994962|
|Semiconductor and...|      292|      172| 0.589041095890411|
|Semiconductor and...|      158|       93|0.5886075949367089|
|Electronic Comput...|      135|       77|0.5703703703703704|
|Computer Systems ...|      920|      519|0.5641304347826087|
|Computer Systems ...|      631|      349|0.5530903328050714|
|   Commercial Banking|      114|       61|0.5350877192982456|
|  Software Publishers|      618|      328|0.5307443365695793|
```

```
+-------------------+---------+
|             sector|app_Count|
+-------------------+---------+
|Custom Computer P...|     4100|
|Colleges, Univers...|     1067|
|Computer Systems ...|      920|
|Computer Systems ...|      631|
|  Software Publishers|      618|
|Engineering Services|      573|
|Other Computer Re...|      510|
|Elementary and Se...|      416|
|General Medical a...|      397|
|Investment Bankin...|      360|
|Offices of Physic...|      341|
|Full-Service Rest...|      303|
|Semiconductor and...|      292|
|Radio and Televis...|      274|
|Administrative Ma...|      212|
|  Electronic Shopping|      165|
|Semiconductor and...|      158|
|Offices of Certif...|      149|
```

```
+-------------------+---------+
|             sector|acc_count|
+-------------------+---------+
|Custom Computer P...|     2156|
|Colleges, Univers...|      690|
|Computer Systems ...|      519|
|Engineering Services|      354|
|Computer Systems ...|      349|
|  Software Publishers|      328|
|Investment Bankin...|      259|
|Other Computer Re...|      244|
|General Medical a...|      235|
|Radio and Televis...|      228|
|Offices of Physic...|      216|
|Elementary and Se...|      190|
|Semiconductor and...|      172|
|Administrative Ma...|      151|
|Full-Service Rest...|      119|
|  Electronic Shopping|      118|
|Offices of Certif...|       94|
|Semiconductor and...|       93|
```