MIS 587 – Business intelligence

H1b Visa petition

**Table of Contents**

[Business Problem 2](#_Toc512986774)

[Data Explanation 2](#_Toc512986775)

[Project Goals 2](#_Toc512986776)

[Solution 2](#_Toc512986777)

[Data Cleaning 2](#_Toc512986778)

[Clustering 3](#_Toc512986779)

[Visualization Using Tableau 5](#_Toc512986780)

[Results 5](#_Toc512986781)

[APPENDIX 6](#_Toc512986782)

# Business Problem

We found this dataset on Kaggle. It involves data on H1B visas of employment-based, non-immigrant visas for temporary foreign workers in the United States. A US employer must offer an international a job and submit a petition for a H1-B visa to the US immigration department for a foreign national to apply for H1-B.

# Data Explanation

We had one dataset which had approximately 3 million records in it. The columns in the dataset consists of the following variables:

* Case status
* Employer Name
* Worksite
* Job Title
* Prevailing Wage
* Full Time
* Year

*Data Set Sample*



# Project Goals

For this project we will provide insights on the wages, number of petitions and top employers that file for H1B.

# Solution

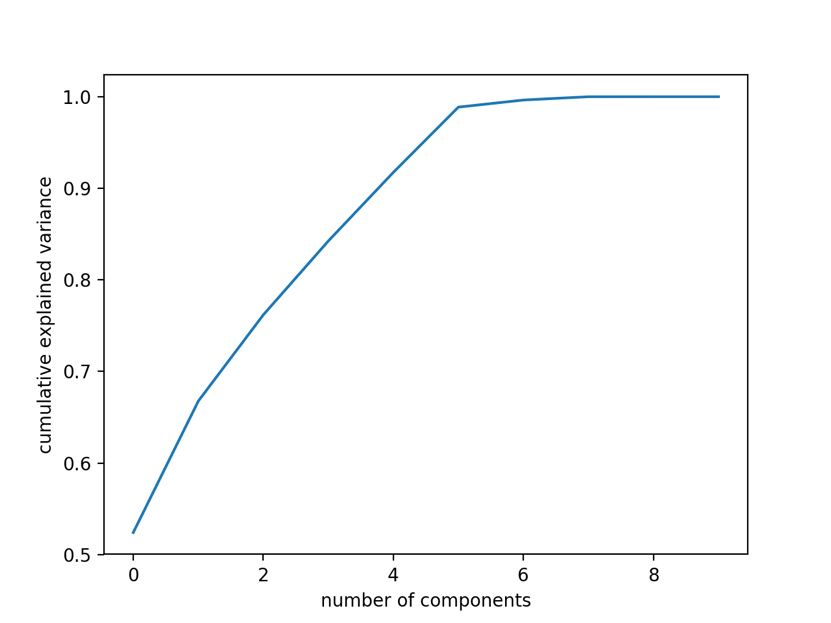
## Data Cleaning

The dataset consisted of many NULL and NA values, so we performed some cleaning on the data using Microsoft Excel.

## Clustering

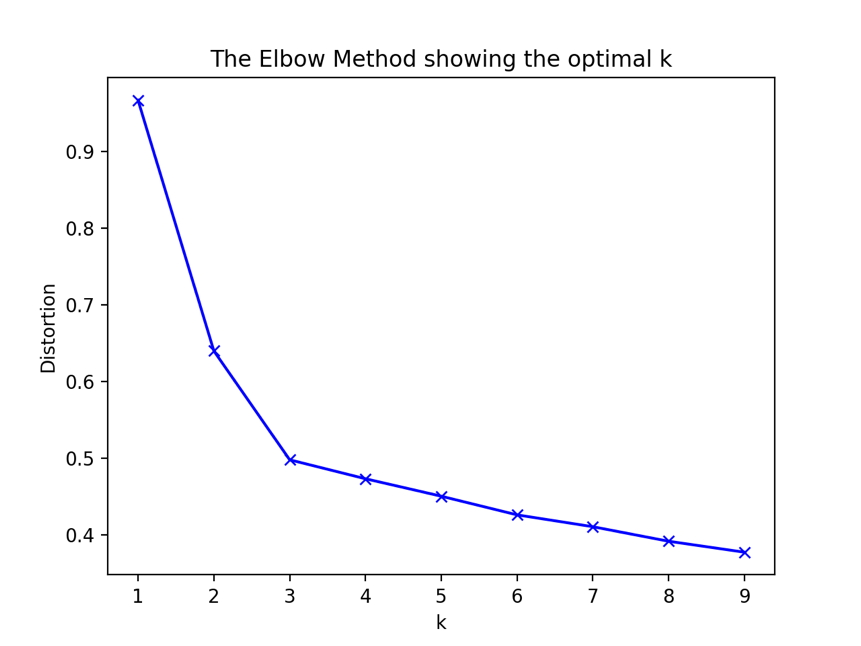
Initially, we converted the categorical data to quantitative data using One-hot encoding and Label Encoding. Once the data was converted to quantitative data we had to perform normalization. The process of normalization brings all the values to the same scale. We had used the Min-Max Scalar to normalize the data and brought it to the same scale. After normalizing the data, we had to perform dimensionality reduction as there were too many columns because many columns were created after the One-hot encoding process. To identify the correct number of dimensions we had to run the PCA multiple times and then had to plot the variance plot to identify the optimal number of dimensions.

*PCA Variance Plot*



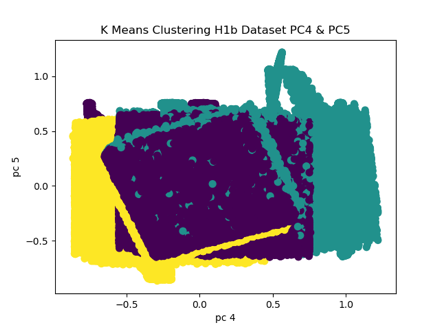
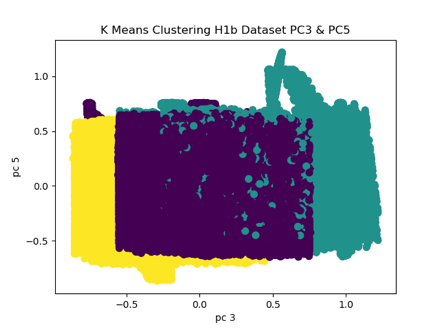
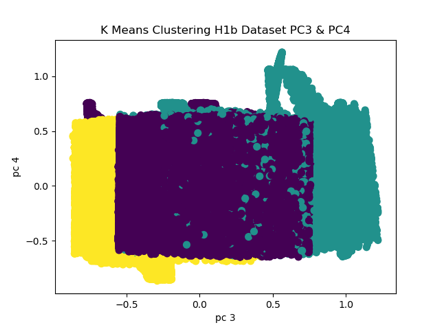
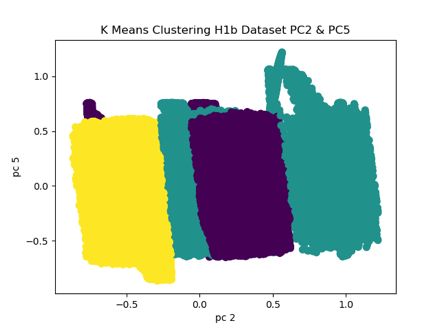
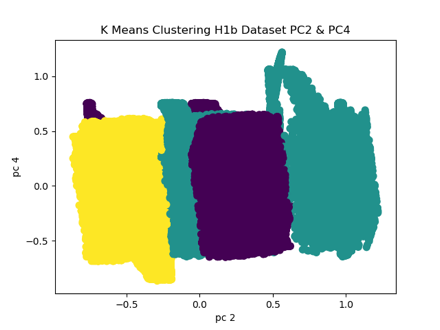
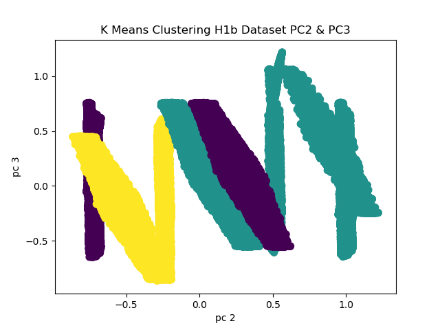
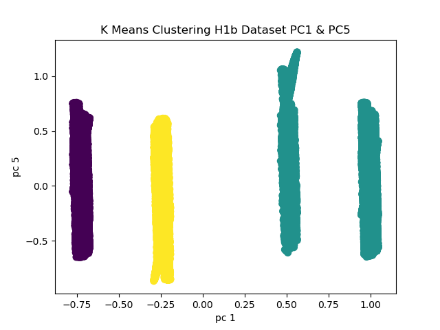
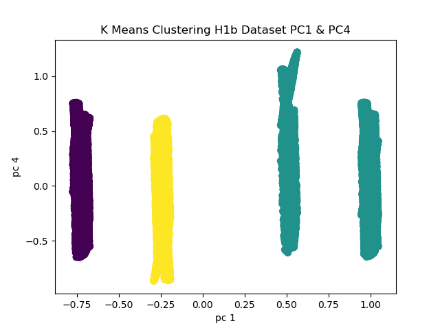
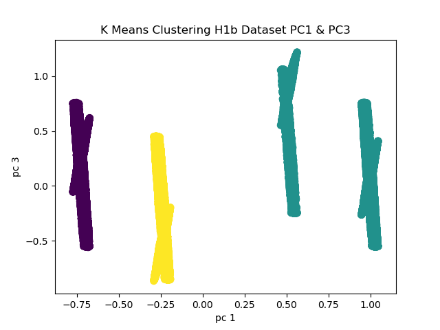
From the above plot we identified that the optimal number of dimensions were 5 as this had the maximum variance. The we derived the 5 PCA components and then we ran K-Means clustering on the PC’s obtained. To find the optimum number of clusters we used the elbow method. The plot for elbow method is below

*K-Means Clustering Elbow Plot*



From the above method we identified the optimal number of clusters to be 3 and then ran the K-Means algorithm to identify the clusters. The cluster visualizations are as below:

*PCA Clustering Visualization*



We used python libraries to run the clustering algorithm. The code for the clustering is attached in the Appendix.

## Visualization Using Tableau

To get better visualizations of the data we merged roles.

We had 10,000 job roles and had to merge them for better visualizations.

CONSULTANT🡪CONSULTANT ROLE

MANAGER 🡪 MANAGER ROLE

PROFESSOR 🡪PROFESSOR

LEAD 🡪 LEAD ROLE

SOFTWARE, DEVELOPER, APPLICATION, PROGRAMMER 🡪 SOFTWARE ROLE

DATA SCIENTIST 🡪 DATA SCIENTIST ROLE

DATA ANALYST, DATA ENGINEER 🡪 DATA ROLE

ANALYST 🡪ANALYST ROLE

# Results

Some of the insights that we had obtained from our visualizations are as follows:

* The state of California has seen the highest number of H1-B petitions till date. It has almost double the number of petitions compared to next highest state Texas.
* The Employer INFOSYS LTD has made the highest number of H1-B petitions till date again making more than double the number of petitions compared to the next highest employer TATA CONSULTANCY SERVICES LIMITED.
* Most of the Petitions fall in the salary range of 60 – 70k USD.
* Washington and California are among the states offering the highest median wages for H1-B jobs.
* There has been an exponential increase in the number of DATA SCIENTIST roles in the last few years.
* Facebook Inc. and Apple Inc. are among the employers that pay the maximum wages for H1-B jobs.
* Software/Tech jobs takes up a major chunk of the total number of H1-B petitions.