

TABLE OF CONTENTS

SR.NO	TOPIC
1)	TITLE
2)	TEAM MEMBERS
3)	ABSTRACT
4)	DATASET SUMMARY
5)	DATA CLEANING
6)	DATA SUMMARY
7)	DATA VISUALIZATION
8)	CONCLUSION
9)	SUGGESTION

TITLE

Python Final Project

***Long Term Occupational Employment
Projection***

Under the Guidance of Prof. Rabih Neochi

Team Members

Sumer Pariani
Siddhesh Dhoot
Raghothama Rao Pranesha
Abhishek Bhogal

ABSTRACT

To help people and organisations make educated decisions about their own careers and organisational programme development, long-term occupational projections for the State of California and its labour market areas are offered for a 10-year time frame. The long-term forecasts are updated every two years. It is well known that decisions about one's career and organisational programme growth directly affect the region's skill/labour market. This information reveals the job market's requirement for certain talents and the possibility for personal career progress for people who acquire those skills. Confounding factors, including job opportunities, entry-level education, work experience, and job training, affect the demand for skills in the labour market. Confounding elements that affect a person's personal development include the sort of neighbourhood they reside in and their line of work.

DATASET SUMMARY

In our project report, we used a dataset with ~15 thousand observations and with 17 variables.

Id: This is an index of the rows variable type int

Area Type: Describes the type of area individual will be working at variable type categorical

Area Name (County Names): Name of the county variable type categorical

Period: The years individual should be actively working

SOC Level: Occupation category divided into 4 types (from 1-4) namely major group, minor group, broad occupation, and detailed occupation

Standard Occupational Classification (SOC) - Subgroup under SOC level variable type categorical

Occupational Title: The designation of the individual

Base Year Employment Estimate: Employee's presently working at the same designation

Projected Year Employment Estimate: Employees that are projected to work at same designation in projected year

Numeric Change : Projected Year Employment Estimate
-Base Year Employment Estimate

Percentage Change: $(\text{Base Year Employment Estimate}) / (\text{Numeric Change}) * 100$

Exits: Exits from that occupation in the given period

Transfers: Transfers from that occupation in the given period

Total Job Openings: Total number of Job Openings occupation in the given period

Median Hourly Wage: Median Hourly Wage of the occupation

Median Annual Wage: Median Annual Wage of the occupation

Entry Level Education: Basic Education needed for the occupation

Work Experience: In order to be eligible for the job, what level of education is needed

Job Training: If the job asks for any kind of Training

DATA CLEANING

In the Dataset, percentage of missing value was 3.87% of the total values. We dealt with different columns for filling the missing values.

The majority of the missing values and cleaning was done for the column Median Hourly Wage, Median Annual Wage, Entry Level Education, Work Experience, Job Training.

For the median hourly wage and media annual wage, we took the mean of the wage values based on the SOC levels (i.e. 1,2,3,4) for respective columns. Furthermore, we replaced all the zero values based on the condition for the SOC levels and if the wage column were zero.

By this method we were able to replace all the zero values in the median annual wage and median hourly wage

For the Entry Level Education null values, we replaced the NaN to 0, after we took the assumption that all the 0 values will be **No information available** as this dataset was for the prediction value and there might be some information which may not have been provided by the employer.

In Job Training we had 3704 NaN and None values which we relaced to 0. After that, we replaced the rubbish data with no training.

For the work Experience Column of null values, we had 4 unique values out of which 2 were None and NaN values. For filling the place holder value, we filled None and NaN to 0, then we filled the 0 values based on the SOC level (1,2,3,4) if SOC level is 1 then we replaced the corresponding 0 value to no experience similarly for SOC 2 we replaced 0s work experience to <5 years and remaining SOC Level 3,4 we replaced corresponding 0 values to greater than equal to 5 years

DATA VISUALIZATION

Our python project analyses several aspects of the present job market and shows a glimpse of how the job market will look ten years from now. Analysis of Long Term employment projection will help every individual in the job market to make informed decisions about What to study, where to find a job, what positions to aim for, what industry to aim for and most importantly, what salary to expect.

Every job in the market is classified under SOC level, I.e. common occupation classification level.

To give a glimpse

1 Total, All Occupations

SOC level 1 are freelancers, people working in less organised sectors, e.g. office staff

2 Installation, Maintenance, and Repair Occupations

SOC level 2 has operators, technicians, Farmers

3 Operations Specialties Managers

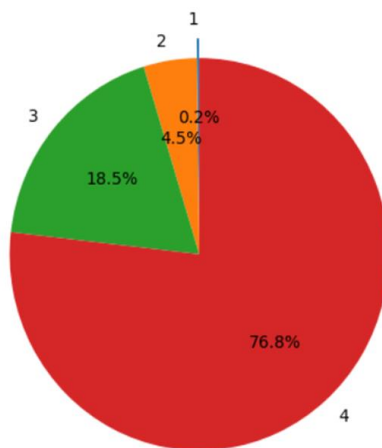
Soc level 3 has Engineers, marketers, managers

4 Chief Executives

SOC level 4 has top executives, top managers, Senior Engineers

Let's first talk about the distribution of jobs these jobs according to their SOC levels in next 10 years

PIE CHART



Pie chart modelling

We grouped the SOC level and took the count of all the SOC values in the dataset. This gave the 4 different parts of the cumulative. This gave us a dictionary of SOC level which in turn would help us to find out portion of individual SOC level in dataset .

Visualization

This pie chart clearly shows that SOC that people working as freelancers or in less organized sectors will be just 0.2% of the whole job market. While Semi Skilled Workers like operators, technicians and farmers will form 4.5% of the entire job market. We see that engineers, marketers and managers will occupy 18.5% of the jobs. At the same time, more than 3/4th of the call will offer employment to managers, Senior engineers and top executives.

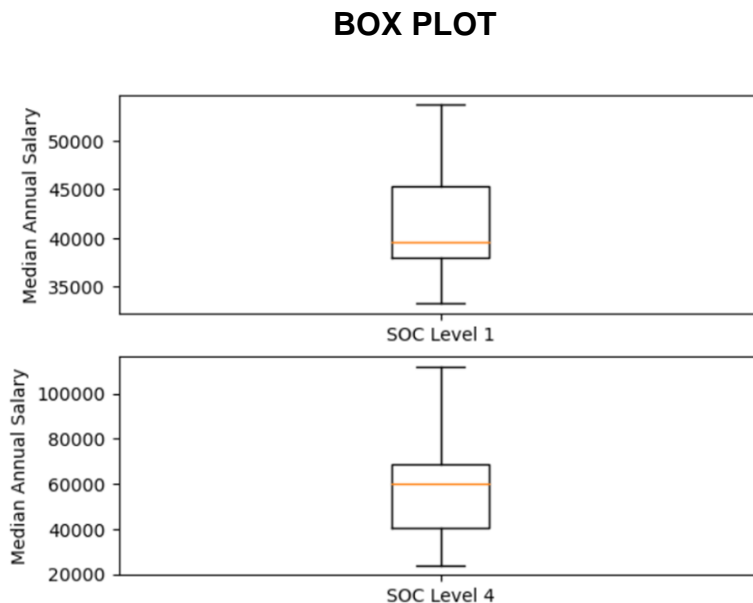
Inference

In the next ten years, the job market will need top executives to take managerial decisions and flourish their Business. This clearly shows that the people studying management, people who have gained sufficient experience as managers, and the people who have experience of

more than five years and have been promoted to senior designations like executives will have a lot of job openings.

Suggestions

This visualization Suggests people should work towards gaining managerial skills and try to work in an organized sector where they could get promoted to senior levels within five years.



Box plot modelling

We found the count of median annual wage of each SOC level. And then created box plot using matplotlib library.

Visualization.

This visualisation clearly shows the salary range an individual would earn while working under a particular SOC level. We can see that jobs in the unorganised sector will have a salary range between 20k to 50k, where 25% of the people working in SOC level 1 will earn below 37.5k, and about 50 % of people working in SOC level 1 will earn under 42k. About 75% of people working under SOC level 1 will earn below 45K. We can see that Executives, managers and senior positions will have a salary range between 25k to 100k, where 25% of the people working in SOC level 4 will earn below 35k, and about 50 % of people working in Soc level 4 will earn under 65k. About 75% of people working under SOC level 1 will make below 75K.

Inference

We can see less variability in salaries for the unorganised sector. This also means that the organised sector's growth potential is shallow. There is a possibility that you can earn nearly

the same as 25% of the managerial jobs. Still, the growth potential under an unorganised sector will not allow you to grow if you keep working in that sector. In contrast, administrative jobs, top executives and senior positions show excellent variability in income. You may earn less in the initial years. Still, the growth opportunities are there, seeing the varied salaries offered to senior designations.

Suggestions-Try to acquire a managerial position even if the salary is low because this sector shows significant variability for its individuals. If you are working in this sector, there is a high chance that you will earn 200% more in the coming years.

LINE GRAPH



Line chart modelling

We grouped SOC level and got the sum of all the values of that SOC level in a dictionary. Then we used matplotlib function to plot the line graph between SOC level and sum of salaries.

Visualization

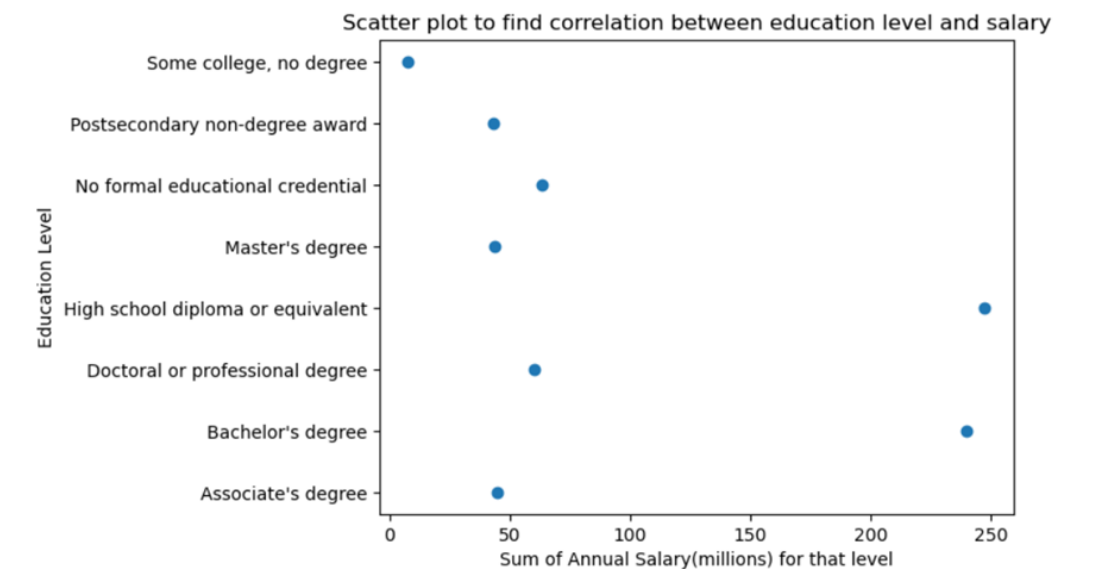
This graph shows the Sum of salaries of people working under different SOC levels. We can see that the Sum of salaries of all individuals working under SOC level 1 is around 1.3 million. In contrast, the Sum of wages for people working under SOC level 2 is about 70 million. The Sum of the salaries of people working under SOC level 3 is about 160 million, and the Sum of wages of people working under SOC level 4 is above 700 million.

Inference

Fewer people will work in unorganised sectors that offer an average salary of 40k in a year. We can see that the difference between people working under SOC level 2 and SOC level 3 is a little, but their sum salaries have a considerable difference. We can also see that his salary will increase exponentially if a junior Engineer gets promoted to a Senior engineer or managerial position. We can infer this from this line graph of the Sum of salaries.

Suggestions If you are working under the SOC level, try to shift to SOC level 2 for a better increment and if you are working under SOC level 3, try to get promoted to SOC level 4 for a better increment.

SCATTER PLOT



Scatter plot modelling

We found the sum of wages for each education level category and plot the on a scatter plot.

Visualization

This scatters plot shows the relationship between entry-level Education and salaries. We can see that individuals with no degree occupy positions under SOC level 1 and thus come under the low-income group. The population of people working with a Diploma or bachelor's degree is very high; therefore, their salaries are exceptionally high. People with diplomas and bachelor's degrees are scattered under SOC levels 2,3 and 4. Fewer people pursue Master's and PhD who Occupy senior technical and managerial positions; thus, their sum salaries are around 50k.

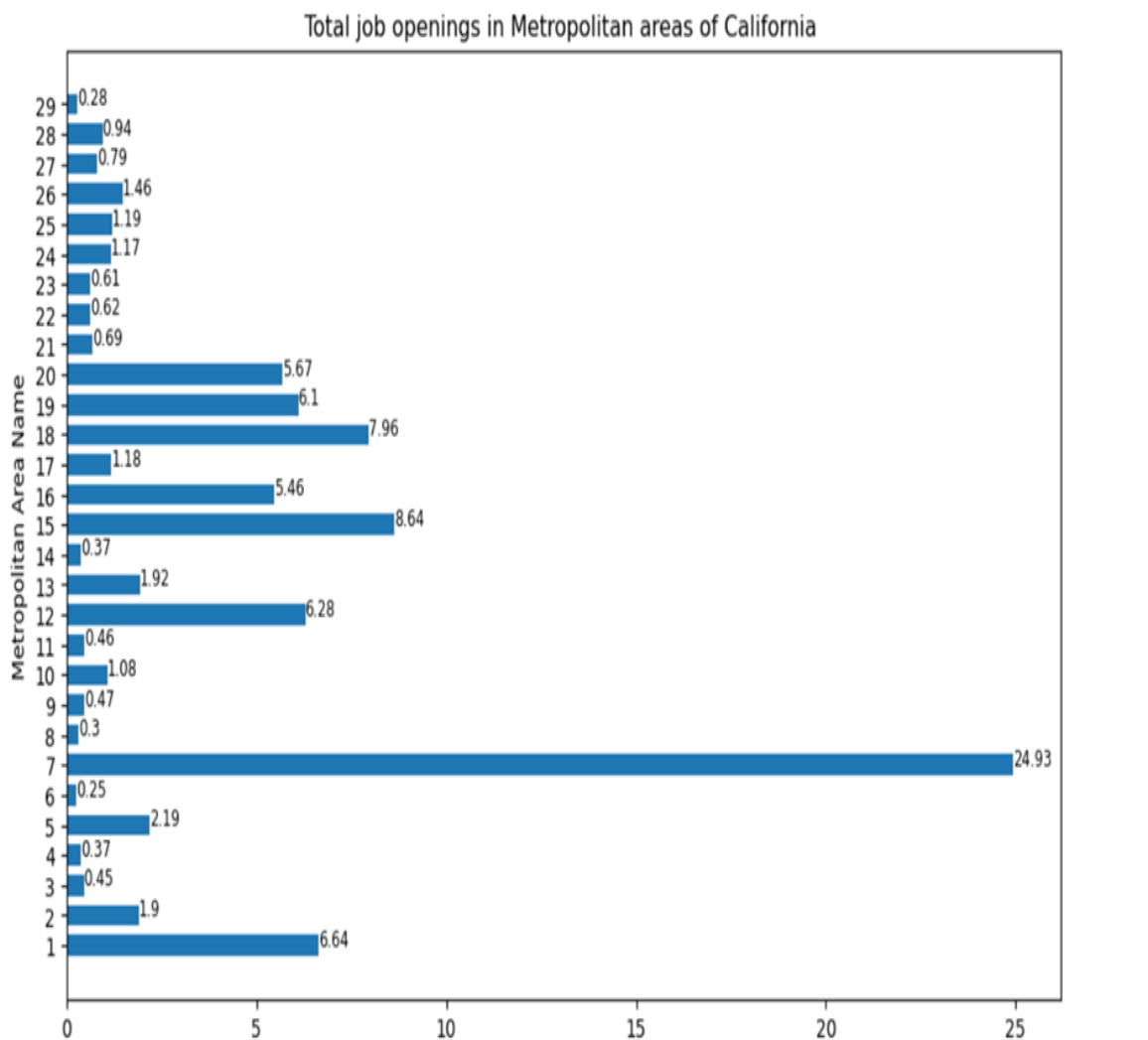
Inference

Few people plan to Secure a senior position during their academic years; thus, we can see that the number of people with master's and PhD is less. Many people come under SOC levels 3 and 4 with just a high School Diploma and bachelor's degree; thus, Education plays a small role in career growth potential, but Work Experience and job training are significant to growth in their career paths.

Suggestions

If you have an opportunity to work and acquire job training in a particular field, grab that opportunity because that will play an important role and receiving a senior designation as soon as possible

BAR GRAPH



Index	Metropolitan Area
1	Anaheim-Santa Ana-Irvine MD (Orange County)
2	Bakersfield MSA (Kern County)
3	Chico MSA (Butte County)
4	El Centro MSA (Imperial County)
5	Fresno MSA (Fresno County)
6	Hanford-Corcoran MSA (Kings County)
7	Los Angeles-Long Beach-Glendale MD (Los Angeles County)
8	Madera MSA (Madera County)
9	Merced MSA (Merced County)
10	Modesto MSA (Stanislaus County)
11	Napa MSA (Napa County)
12	Oakland-Berkeley-Livermore MD (Alameda and Contra Costa Counties)
13	Oxnard-Thousand Oaks-Ventura MSA (Ventura County)
14	Redding MSA (Shasta County)
15	Riverside-San Bernardino-Ontario MSA (Riverside and San Bernardino Counties)
16	Sacramento-Roseville-Folsom MSA (El Dorado, Placer, Sacramento, and Yolo Counties)
17	Salinas MSA (Monterey County)
18	San Diego-Chula Vista-Carlsbad MSA (San Diego County)
19	San Francisco-San Mateo-Redwood City MD (San Francisco and San Mateo Counties)
20	San Jose-Sunnyvale-Santa Clara MSA (San Benito and Santa Clara Counties)
21	San Luis Obispo-Paso Robles MSA (San Luis Obispo County)
22	San Rafael MD (Marin County)
23	Santa Cruz-Watsonville MSA (Santa Cruz County)
24	Santa Maria-Santa Barbara MSA (Santa Barbara County)
25	Santa Rosa-Petaluma MSA (Sonoma County)
26	Stockton MSA (San Joaquin County)
27	Vallejo MSA (Solano County)
28	Visalia MSA (Tulare County)
29	Yuba City MSA (Sutter and Yuba Counties)

Bar graph modelling

we took the sum of total job openings in the dataset, grouped the job opening according to county name and created a dictionary. Then we used matplotlib to create a horizontal bar graph to find the relation between total job opening and county names

Visualization

These will graph people finding job openings in different locations. Based on job opening places, counties are divided into categories.

A cold category where job openings are between 0.28 to 2.9 million

A warm type where job openings are between 5 million to 7 million

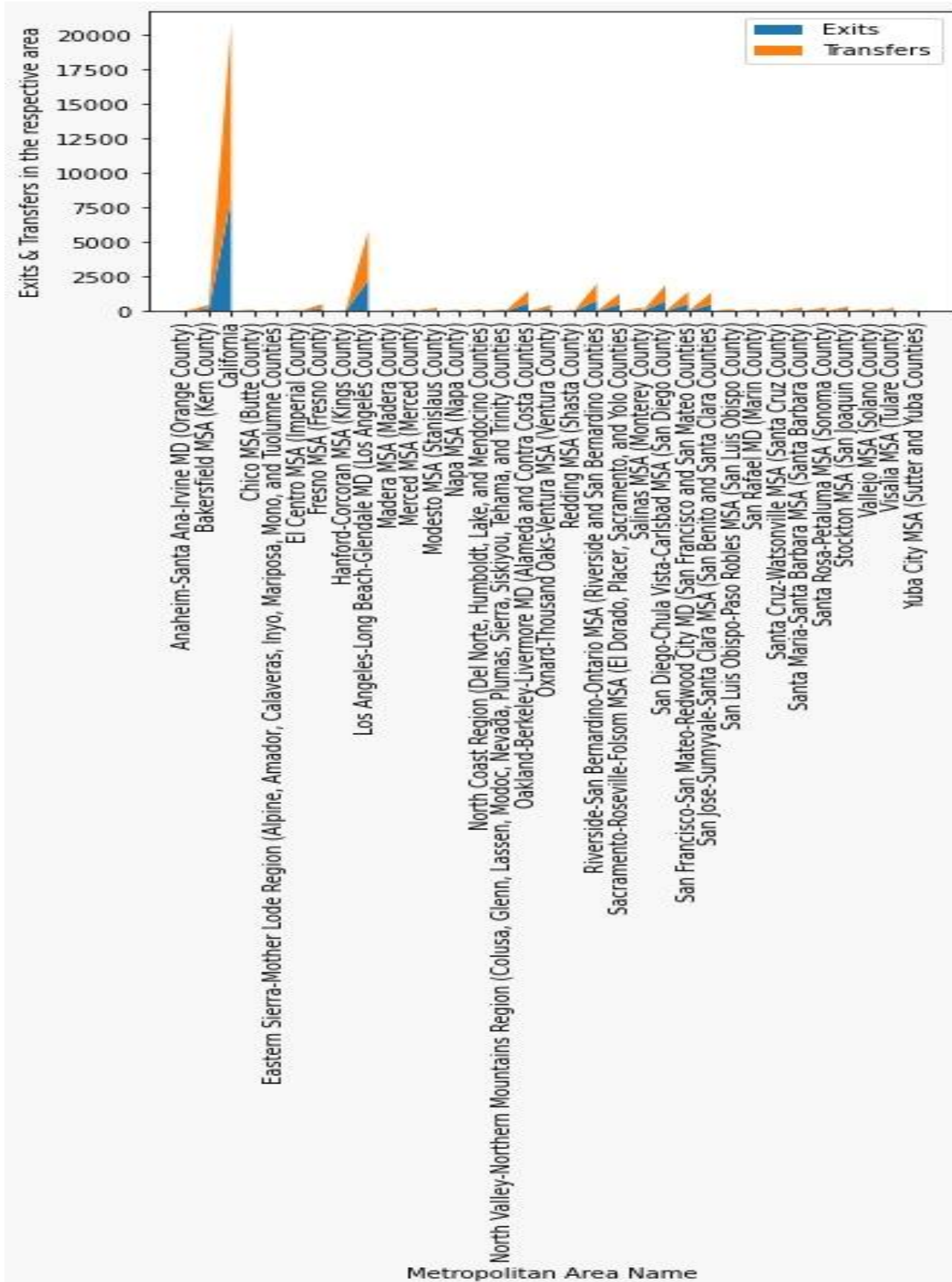
A hot variety where job opening is above 10 million. Which include Bernardino and Los Angeles

Inference

All Locations would have a job opening for every SOC level category but staying near Warm and HOT places increases the securing of a job exponentially.

Suggestions Suppose you want periodically switches company warm and hot categories. In that case, counties will be the places for you to work because you can find great opportunities there as there are a lot of job openings in those areas

Stacked plot



Stacked plot modelling

We used matplotlib. Stacked plot to create a stacked plot between county and the transfers and exits the county has witnessed.

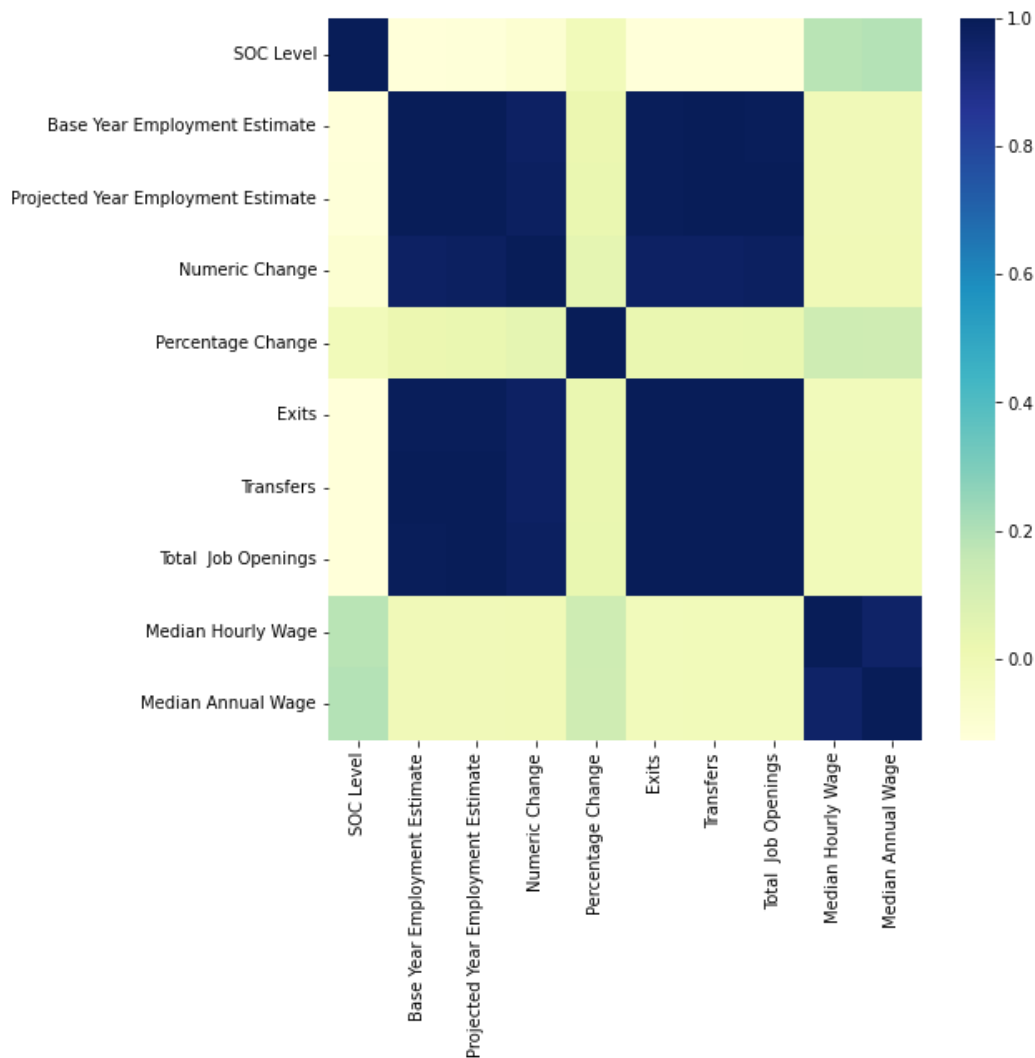
Visualization

This graph shows that the number of exits and transfers are high for Chico MSA county. And other places of cold and warm categories also witness average number of exits and transfers.

Inference

Individual working in cold and warm categories place witness high number of exits and move towards hot category counties because the greater number of job opening at that location.

Heat map



A Heat map is to find dependency of one variable with other. In our heat map we observe that median hourly wage and median annual wage columns have a high co relation with SOC level as we have used SOC level to fill the missing values in these two columns which were later used for visualizations.

CONCLUSION

In this Data Visualization Project, we understood the process of data cleansing, data gathering and selecting the appropriate dataset for data analysis. Furthermore, we were also able to understand the process of data visualization using different function, different libraries, plotting methods.

With the help of 7 graphs, we can deduce that

In Next 10 years the focus will change from technology to management side. There will be ample of job opening in the management sector. There is a possibility that even after becoming a manager you still earn the same as any person working under unorganised sector but the variability of the salaries in organised sector is great if you are working in management area there is potential to go beyond 150k annually, whereas variability is less under unorganised sector. We also came to know that people working under SOC level 1 try to switch to SOC level 2 to exponentially increase their earnings and the same is possible for SOC level 3 and SOC level 4. We also got to know that very less people plan their career when they are studying which is shown by less number of people attending Masters and PHD courses and they depend more on work experience and job training to grow in their career. The job openings also depend on the location of the company where we see places are divided into cold, warm and hot categories according to job openings and we also say that places under cold and warm categories see high number of exits and transfers portraying people are shifting from cold and warm places to hot categories counties.

Suggestions

According to the visualizations it is suggested that

An Individual should learn managerial skills

And grab opportunities to gain work experience and job training rather than depending upon courses

Find jobs in categories coming under Hot categories

And try to switch their jobs to get immediate promotions to next SOC level for having a financial healthy and secure job in state of California.

Data set Suggestion

It would be of great help if we could have added a column which shows the population of people who in the first year of corporate life to see yearly influx of people under each SOC level.