JobHarvestPro: A Data-Driven Approach to Job Search

Introduction

The job market is constantly changing, and it can be difficult for job seekers to keep up with the latest trends. However, with the help of data-driven tools, job seekers can gain a better understanding of the market and make more informed decisions about their careers.

This paper presents JobHarvestPro, a data-driven approach to job search. JobHarvestPro uses data from the SimplyHired website to provide job seekers with insights into the job market, such as the most popular job titles, the highest-paying jobs, and the most in-demand skills. JobHarvestPro also provides visualizations that help job seekers to understand the data and make informed decisions about their careers.

Data Collection

JobHarvestPro collects data from the SimplyHired website using the Python library BeautifulSoup. The data that is collected includes the job title, company, location, job type, salary, posted on, job qualification, job description, job link, job site, keyword, posted on date, min salary, max salary, and rating.

We had done a series of extracting data from several websites. We started off with LinkedIn website. We were successfull in getting some data, that is around 1000 rows of data. We quickly learned that there is some issues with this data. If we needed more data we had to hit the authentication wall, which we did not want as this approach is bad. Apart from that the data we had collected had less salary information which was one of the main data points for creating visualizations. Along with this there was another roadblock where in we could not split the data with exact location, state and country. So we decided we would not move ahead with this website. We moved on to trying with another website Indeed.com. There were similar issues as with LinkedIn. Along with that we observed with some keywords we had very less data and for few we could not go past page 36 which was a roadblock for us. We decided to then move ahead with another website SimplyHired.com. Here we got all the information needed to move ahead with the tasks.

Phase1 Extraction:

The selections we chose were as follows.

job_title

company

location

job_type

salary

posted_on

job_qualification

job_description

job_link

job_site

keyword

posted_on_date

min_salary

max_salary

rating

Extracted data with 4 keywords.

1)Software Engineer

2)AI/ML Engineer

3)Data Engineer

4)SAP

Data Transformation

Phase2

The data that is collected from SimplyHired is transformed in several ways. First, the null values are removed. Second, the minimum and maximum salary are separated into two columns. Third, unwanted columns are removed. Finally, the data is formatted in a way that is easy to use for visualizations.

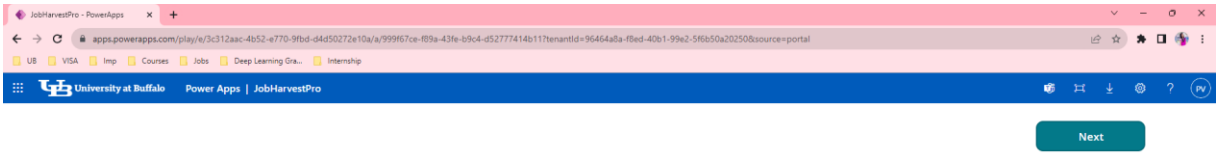
Visualizations

JobHarvestPro uses a variety of visualizations to help job seekers understand the data. These visualizations include cards, slicers, line charts, stacked column charts, donut charts, treemaps, filled maps, clustered bar charts, and stacked column charts.

Phase3

We then created thoughtfull visualisations and also created an app with the data we had using Powerapps.

Our home screen consists of our Job title and our company name APPMANDI LLC. We have included navigation buttons in each page. Each Page has a slice for all the 4 keywords to have a clear distinction for each keywords and information associated with it.



Page 1 consists of the following visuals.

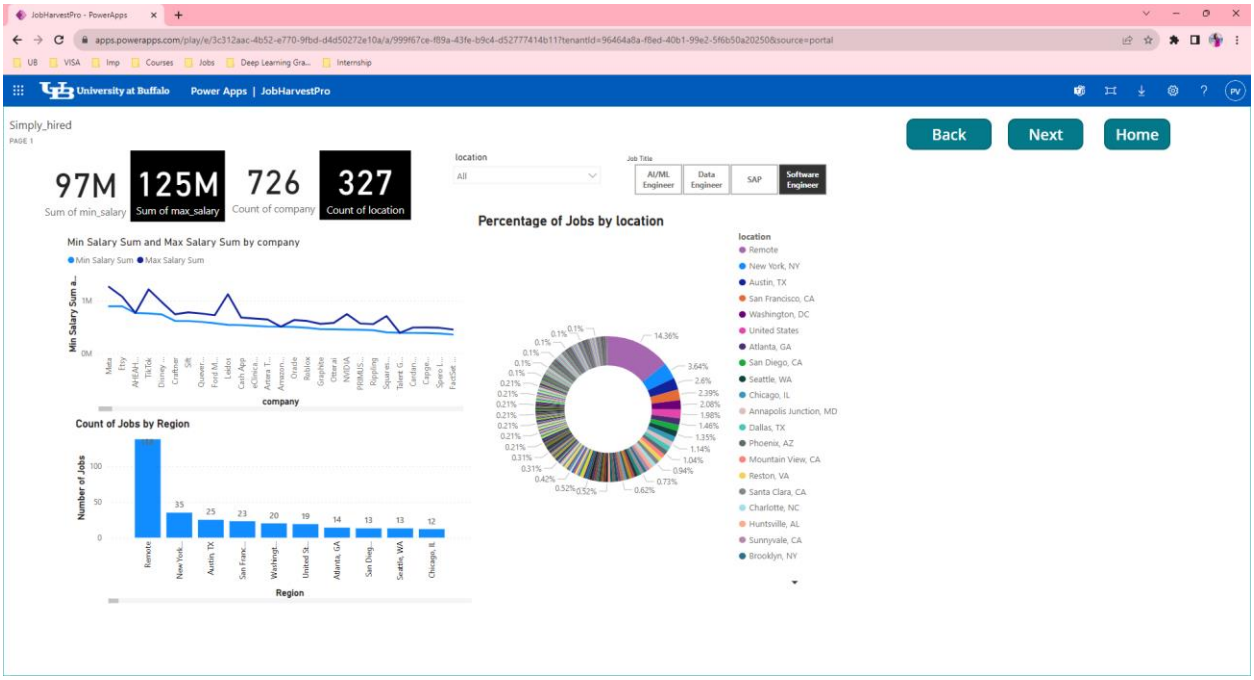
1)Card - This specifoes total minimum salary, total maximum salary, count of total company per kyword and count of the location.

2)Slicer - Here we have 2 slicers. One for location. As we select the location the information associated to a particular location is displayed. Another slicer consists of 4 keywords.

3)Line Chart - This graph shows minimum and maximum salary per company.

4)Stacked column chart - This graph shows the total number of jobs per region.

5)Donut chart - This shows the percentage of jobs per location.



Page 2 consists of the following visuals.

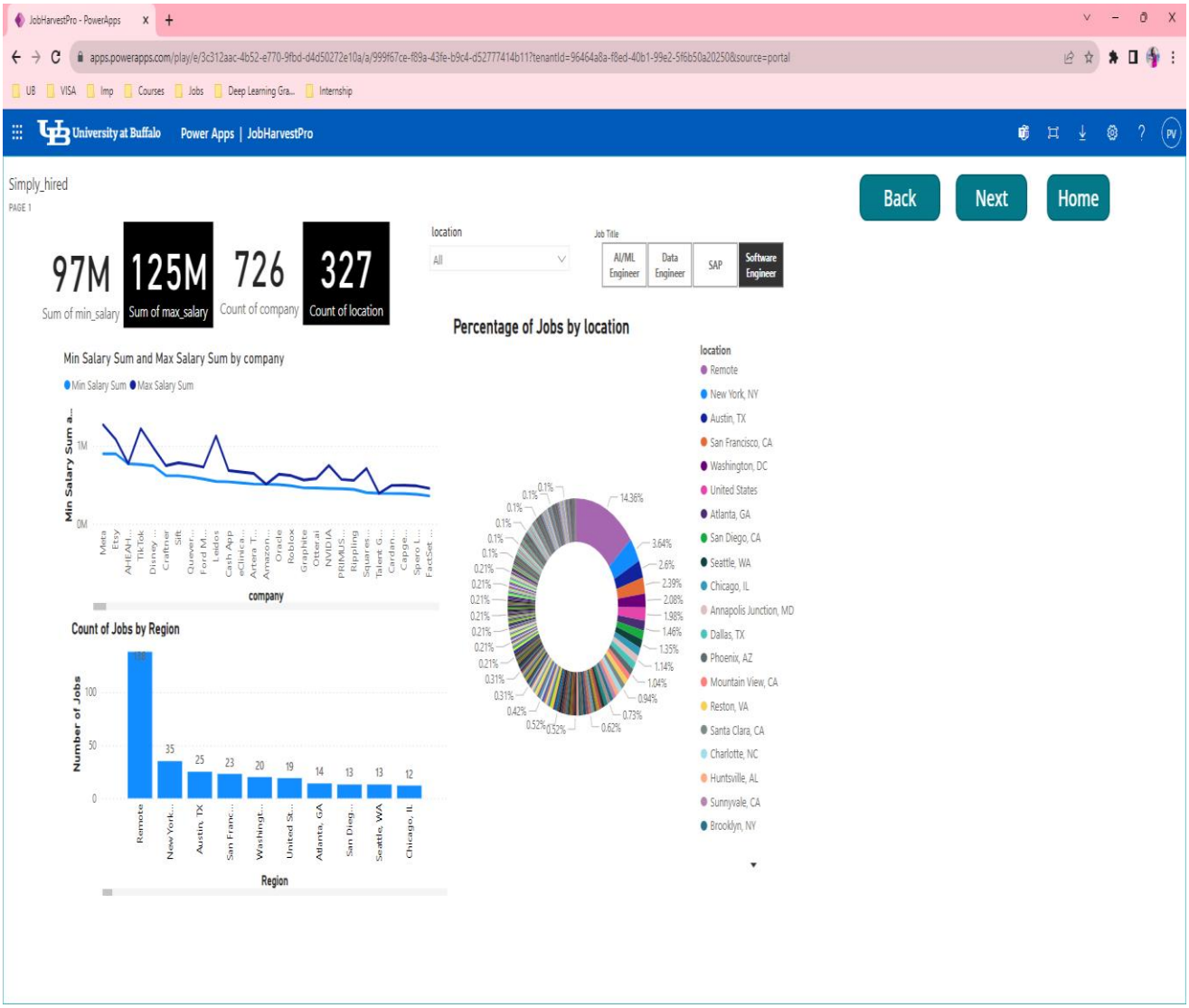
1)Line Chart - displays the total number of jobs posted per day.

2)Treemap - Displays count of the company per location.

3)Slicer - Here we have 2 slicers. One for location. As we select the location the information associated to a particular location is displayed. Another slicer consists of 4 keywords.

4)Donut chart - This shows the percentage of job type.

5)Filled map - Highlights the region where the data is collected.

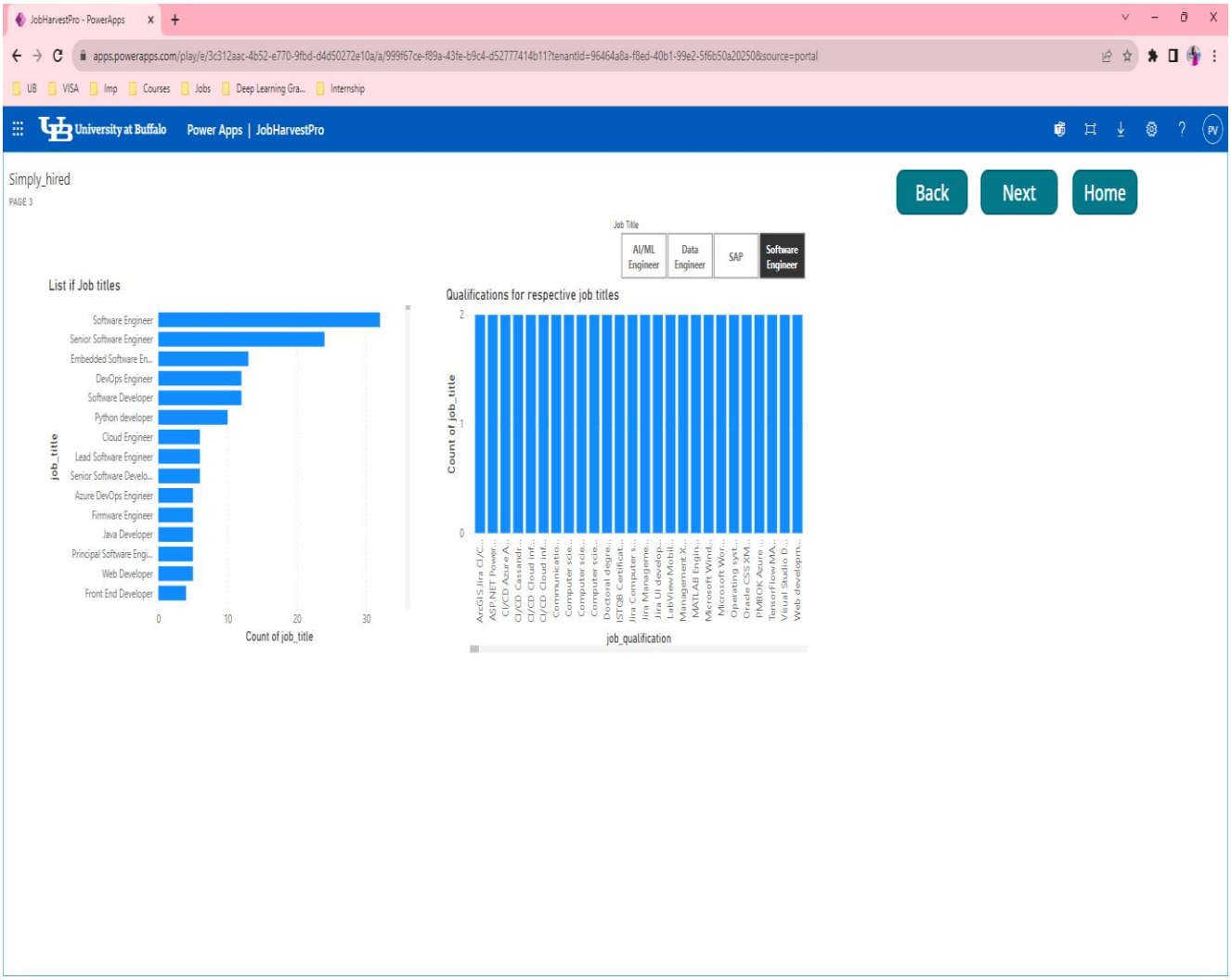


Page 3 consists of the following visuals.

1)Clustered bar chart - Shows count of each job title.

2)Stacked column chart - The job title in the above visual shows the corresponding qualification for each in this section

3)Slicer - Here we have 1 slicer.Slicer consists of 4 keywords.

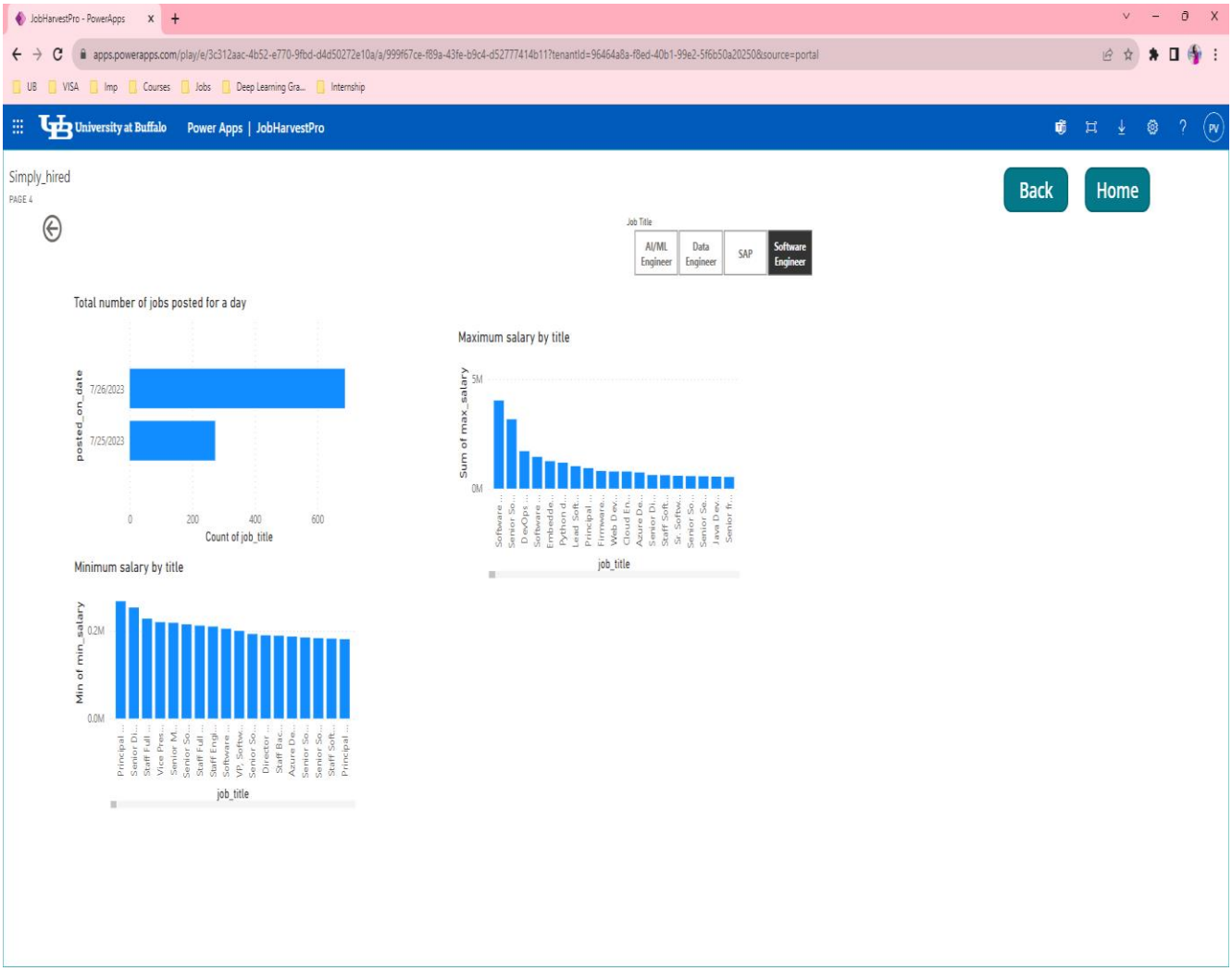


Page 4 consists of the of the following visuals.

1)Clustered bar chart - Shows the number of jobs posted each day.

2)Stacked column chart - There are 2 charts here one for maximum salary and one for minimum salary. Each shows the total in their respective sections with respect to the job title.

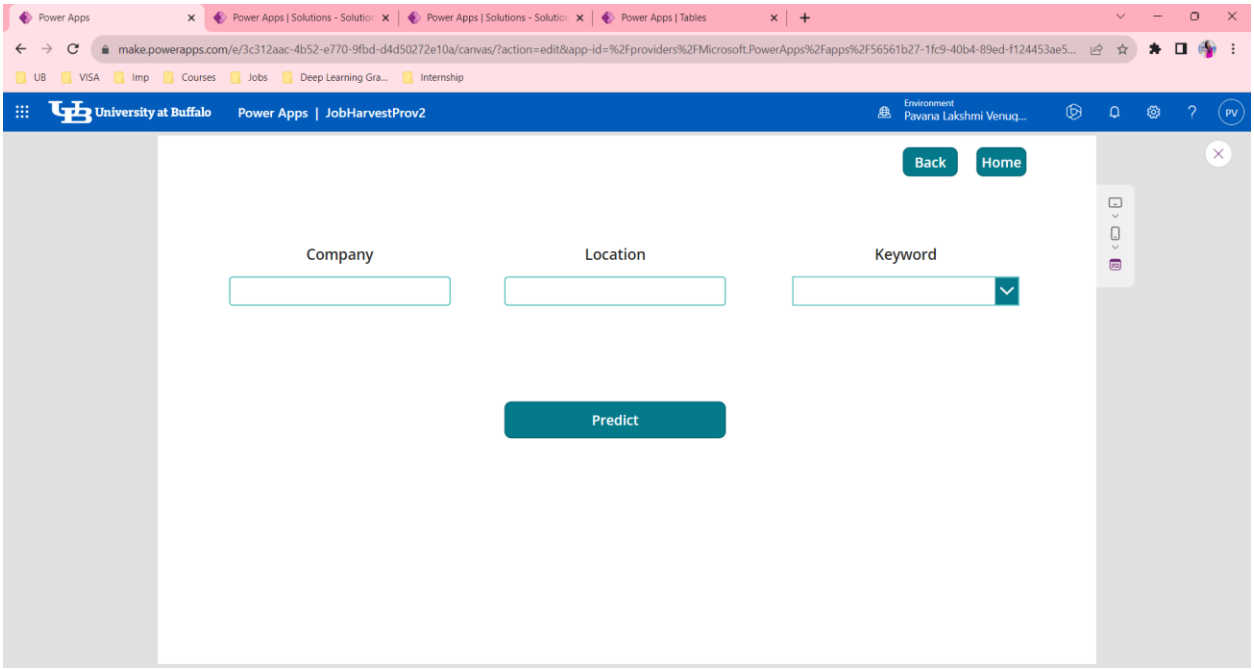
3)Slicer - Here we have 1 slicer.Slicer consists of 4 keywords.

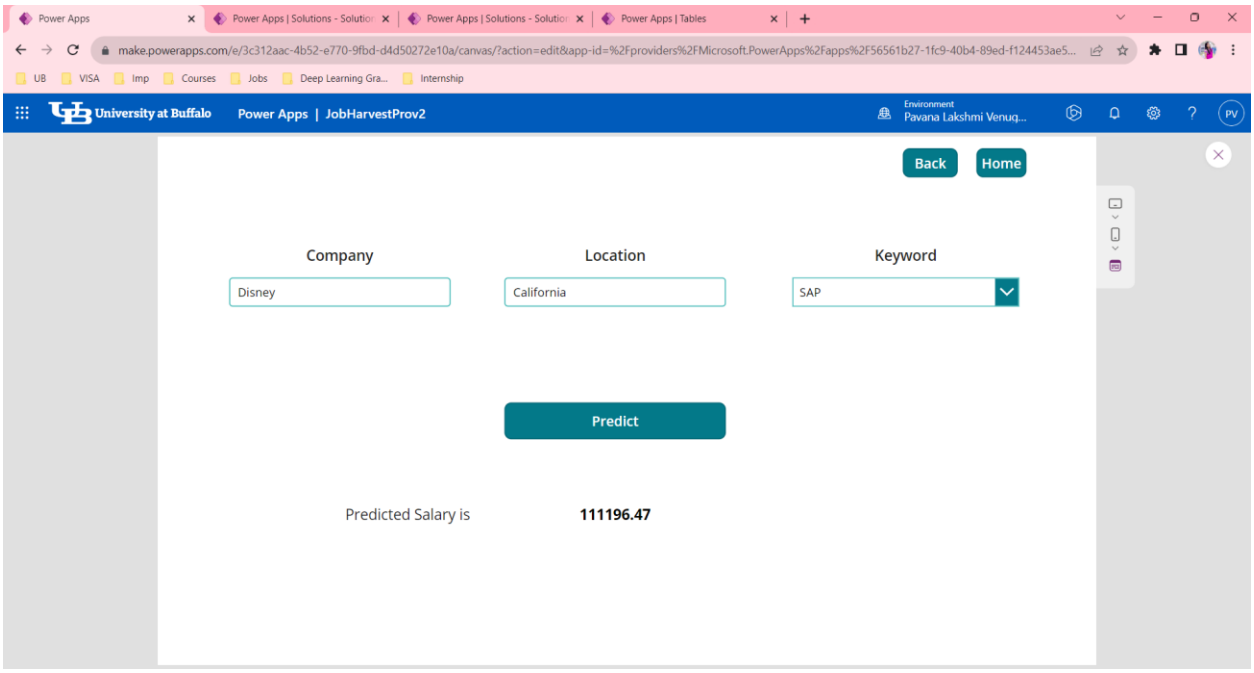


Prediction

JobHarvestPro also uses a prediction model to help job seekers predict the salary. The prediction model is trained on a dataset of historical job postings and job seekers' profiles. The model uses the data to predict the probability of the salary for each job.

Phase4





Conclusion

JobHarvestPro is a data-driven approach to job search that can help job seekers to gain a better understanding of the job market and make more informed decisions about their careers. The tool provides insights into the job market, visualizations that help job seekers to understand the data, and a prediction model that can help job seekers predict their expected salary.

References

- https://www.simplyhired.com/