Project Title
Data Science employment opportunities in Australia

Team Members
1: Anam Khalid
2: John Antony
3: Nishant Patel

Project Description/Outline
Data Science opportunities in Australia

Research Questions to Answer
Anam:
1: Find the top companies with the highest number of job opportunities
2: Find the top companies with greatest estimate low, base and high salary
3: Display top 3 job titles with highest number of jobs using box plot
4: Find the unique job titles and extract top 5 and bottom 5 job titles in terms of salary
5: Analyze which companies have most and least satisfied employees.
6: Find the most in demand technical skills required by the companies

Nishant:
7: Find the state with the highest and lowest number of job opportunities and create bar chart for state vs job opportunities
8: Find the popular company type? and find company-private and company-public for job opportunities. create pie charts for company-private job opportunities vs company-public job opportunities.
9: Compare the size of company and estimate base salary using the bin method

John:
10: What are the Top 10 and Bottom 10 Companies based on Company Rating, Company Career Opportunities, Compensation and Benefits, Company Culture and Values and Company Work Life Balance"?
11: Is there a strong linear regression between:
   ● Company Rating vs Estimate Base Salary
   ● Company Rating vs Compensation and Benefits
   What conclusion can we make from this analysis?
12: Can we use Gmaps API to portray which cities have the highest job opportunities, highest max salary and highest average salary? And what does this explain about the data set?
13: What is the average estimated salary for a metro and regional location? Also, portray top

Datasets to Be Used

https://www.kaggle.com/datasets/nadzmiagthomas/australia-data-science-jobs

Rough Breakdowns of Task

1: Creating files and pushing to github

2: Importing Libraries

3: Reading through the csv file

4: Making a dataframe

5: Data cleaning

       5.1: Dropping null values

       5.2: Find and remove any unwanted special characters

       5.3: Delete the unnecessary columns

       5.4: Make a new clean dataframe

Anam:

6: Finding and displaying top 6 companies with maximum number of jobs using value_counts() and iloc()   methods.

7: Use groupby(), sort_values() and iloc() method to find and display the companies with the greatest average of low, base and high salary.

8: Find out the maximum number of unique job titles and then display salaries of top 3 job titles using box plot

9: Use groupby() method to find out the top 5 and bottom 5 job titles in terms of salary

10: Calculate the mean of the four columns relevant to the employee satisfaction, find the mean of employee satisfaction and display top ten and bottom ten companies in terms of employee satisfaction using the iloc method.

11: Using sum(), sort_values() and iloc method, find the top ten in-demand skills by the companies.

Nishant:

12: We use a group by method between State and Job Title and Find out which states have higher data job opportunities and based on that analysis we create a bar graph between state vs number of job opportunities.

13: By using the Loc method we extract information about two job types (company-public and company-private). We merge hospital type into public type and find how many job opportunities for company-public and company-private. Based on that analysis we create pie charts for company-private vs company-public.

14: using as type() we remove the string from the Company size column and then we use the bin method for company estimate salaries and we use the bin method to find the size of the company.

15: we create a bar graph between company type vs estimate base salary.

John:

16. Find top 10 and bottom 10 companies in terms of company ratings and employee satisfaction by setting conditions over 3.5 and less than 4 respectively. For this we created conditions for each classification and groupby the company with the mean of all employee satisfaction measures.

17. Calculate correlation between Company rating and Estimated Base Salary by using linear regression. First identify the correlation by using st.pearsonr then plot it in a scatter plot with a linear regression line. Perform the same steps with Company rating vs Compensation and Benefits.

18. Using Gmaps API to create a heatmap that we can use to visualize which locations have more employment opportunities and highest estimated salary when it comes to data science.

- · Create a DataFrame with Max salary, Average Salary and Count of Jobs and call this summary_df.

- · Set up json response and base url to make a geocode API request.

- · Identify list of locations, and list for lat/long. Then run a for loop to extract all lat and long coordinates for each location.

- · Create a DataFrame with job location, lat, and long. Merge this DF with summary_df.

- · Using the template add job location, max salary and number of job opportunities to the heatmap and store the DataFrame Row.

- · Populate heat map layer for all job locations and add a marker layer. Finally display the figure.

19. Define metro and regional by using conditional statements and find the number of opportunities in metro vs regional.

- · Add a new column called "location type" to classify location as "regional" and metro and merge them into a DataFrame.

- · Calculate average salary based on location type, metro locations and regional locations.

- · Create a bar chart to display Highest Average Estimated Base Salary Per Job Location

- · Create a bar chart to display Average Estimated Base Salary Per Job type

· Create bar charts to display top 10 locations for data science jobs (Overall, Metro and Regional)