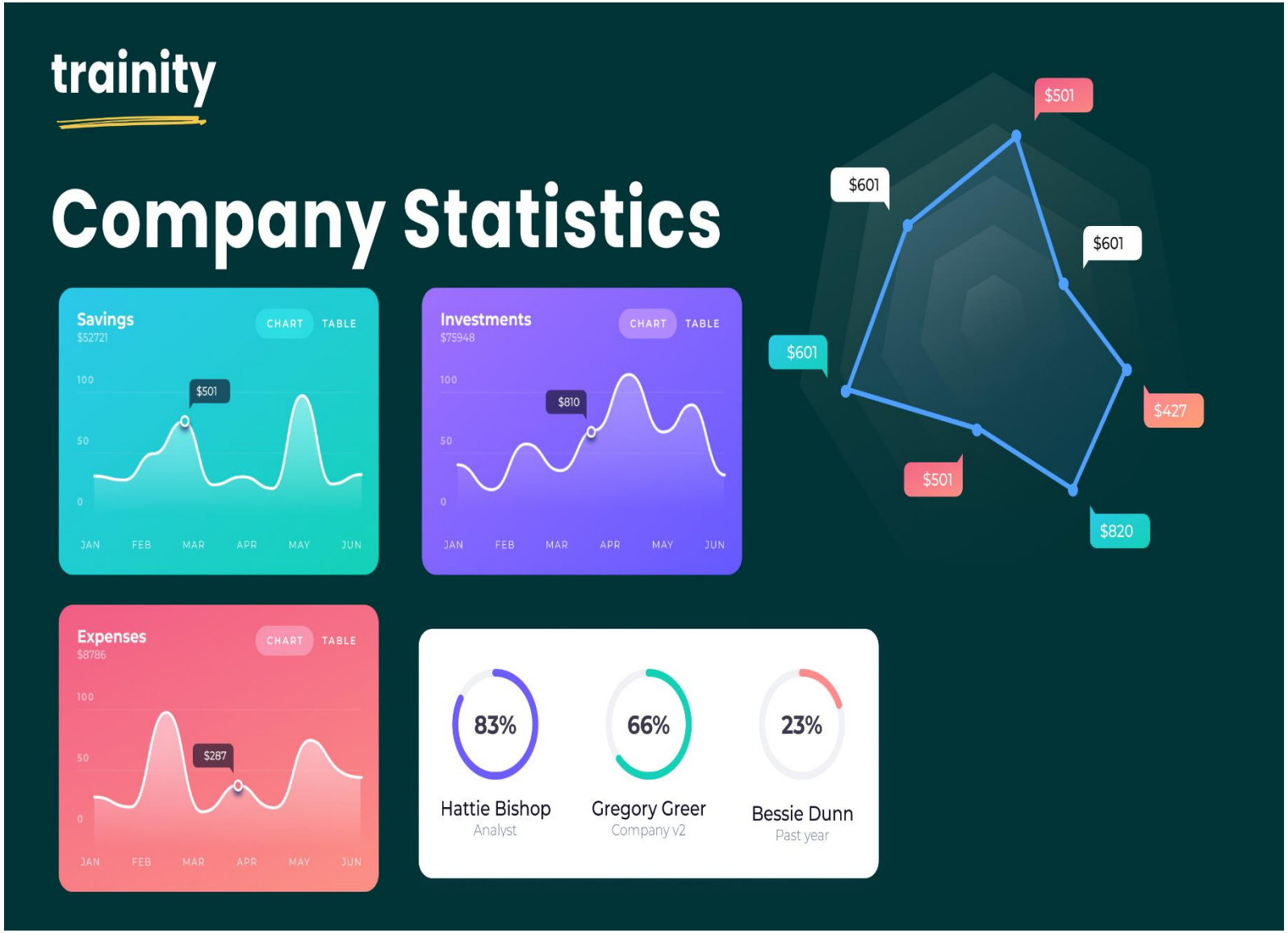
Hiring Process Analytics

Name : Aniket Ashok Ubale
Class : B. Tech
Year : 3rd
Branch : Artificial Intelligence And Data Science
Project : Hiring Process Analytics
Date : 27-04-2023



Project description

Imagine you're a data analyst at a multinational company like Google. Your task is to analyze the company's hiring process data and draw meaningful insights from it. The hiring process is a crucial function of any company, and understanding trends such as the number of rejections, interviews, job types, and vacancies can provide valuable insights for the hiring department.
As a data analyst, you'll be given a dataset containing records of previous hires. Your job is to analyze this data and answer certain questions that can help the company improve its hiring process.

Here's what you'll be doing:
Handling Missing Data: Check if there are any missing values in the dataset. If there are, decide on the best strategy to handle them.
1.Clubbing Columns: If there are columns with multiple categories that can be combined, do so to simplify your analysis.
2.Outlier Detection: Check for outliers in the dataset that may skew your analysis.
3.Removing Outliers: Decide on the best strategy to handle outliers. This could be removing them, replacing them, or leaving them as is, depending on the situation.
4.Data Summary: After cleaning and preparing your data, summarize your findings. This could involve calculating averages, medians, or other statistical measures. It could also involve creating visualizations to better understand the data.
Remember, the goal of this project is to use your knowledge of statistics and Excel to draw meaningful conclusions about the company's hiring process. Your insights could potentially help the company improve its hiring process and make better hiring decisions in the future.

Data Analytics Tasks:
After downloading the dataset, use Excel to answer the below questions:

A. Hiring Analysis: The hiring process involves bringing new individuals into the organization for various roles.
Your Task: Determine the gender distribution of hires. How many males and females have been hired by the company?
B. Salary Analysis: The average salary is calculated by adding up the salaries of a group of employees and then dividing the total by the number of employees.
Your Task: What is the average salary offered by this company? Use Excel functions to calculate this.
C. Salary Distribution: Class intervals represent ranges of values, in this case, salary ranges. The class interval is the difference between the upper and lower limits of a class.
Your Task: Create class intervals for the salaries in the company. This will help you understand the salary distribution.
D. Departmental Analysis: Visualizing data through charts and plots is a crucial part of data analysis.
Your Task: Use a pie chart, bar graph, or any other suitable visualization to show the proportion of people working in different departments.
E. Position Tier Analysis: Different positions within a company often have different tiers or levels.
Your Task: Use a chart or graph to represent the different position tiers within the company. This will help you understand the distribution of positions across different tiers.

Hiring Process Analytics

- Hiring process is the fundamental and the most important function of a company. Here, the MNCs get to know about the major underlying trends about the hiring process.
- Trends such as- number of rejections, number of interviews, types of jobs, vacancies etc. are important for a company to analyse before hiring freshers or any other individual. Thus, making an opportunity for a Data Analyst job here too!

Project – Hiring Process Analytics

Description-

- This Hiring Process Analytics project is all about analyzing a data, do data cleaning draw insights from a data.
- We will perform Exploratory Data Analysis (EDA) where we will try to checking a dataset , checking outlier ,filling a null values
- For EDA we can use Microsoft excel or python labraries (jupyter notebook.
- Here my task is to analyze the dataset provided by the company to extract the usefulinsights and draw conclusions about the company’s recruitment process using statisticalanalysis in Excel or in Google analytics

Approach-

- First we will be performing our analysis on jupyter notebook using various in-built python libraries such as pandas ,numpy, matplotlib,seaborn etc
- We use EDA understanding columns and rows, identifying missing values,handling missing values, checking outliers ,removing outlier.
- we will understand the various columns, the data it contains and their characteristics. Then we will look for duplicates in the data set and if any we will remove it through remove duplicate inbuilt function of python.
- I have used **jupyter notebook** instead of excel or google sheets because I had already knowledge about datasets, python and jupyter notebook tool which I had learn from college.

Software-

- python
- Jupyter notebook

Insights-

- I have done the analysis on the provided dataset / data as per the questions asked and provided the necessary insights and tried to plot the necessary charts /graphs as per requirement and my understanding. This project is helped me in better understanding the process of Exploratory Data Analysis (EDA).
- statistical analysis plays very important role in procces analytics.
- Performing EDA on the data set of HR has helped me to understand the basics steps involved in Exploratory Data Analysis like cleaning of data and deriving inference from the data by performing various statistical analysis.

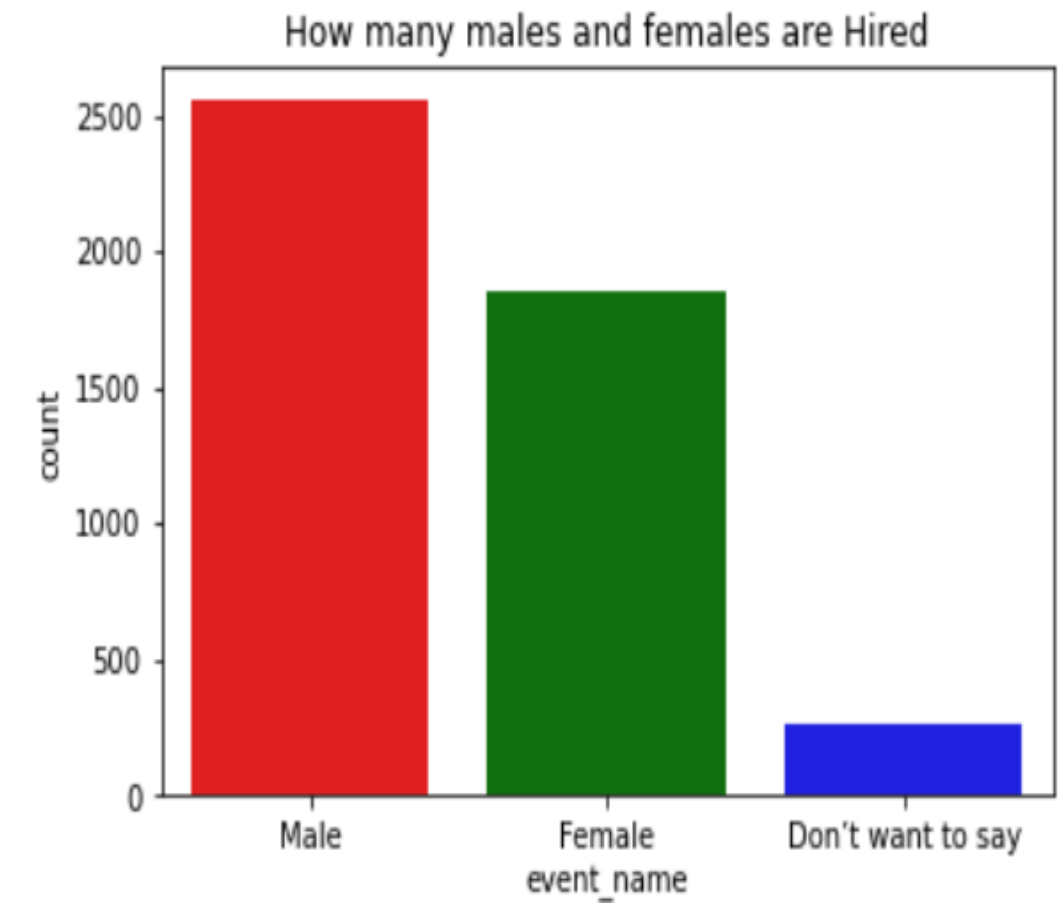
Dataset-

- Table features
- Application_id
- Interview taken on
- Status
- Event_name
- Department
- Post name
- Offered salary

Hiring: Process of intaking of people into an organization for different kinds of positions.
Your task: How many males and females are Hired ?

- Code used-

```
# Filtering males and females hired.
abc=data[data["Status"]=="Hired"]  # only hired
plt.title("How many males and females are Hired") # title of graph
p=["Red","green","blue"]  # for color
sns.countplot(x =abc.event_name, data = data,palette=p) # for count
plt.show()    # for showing graph
```



Average Salary: Adding all the salaries for a select group of employees and then dividing the sum by the number of employees in the group.
Your task: What is the average salary offered in this company ?

- code used-

```
abc=data.groupby(["Post_Name"]).Offered_Salary.agg(["mean"])
abc.rename(columns={"mean":"Average_Salary"},inplace=True)
Abc
```

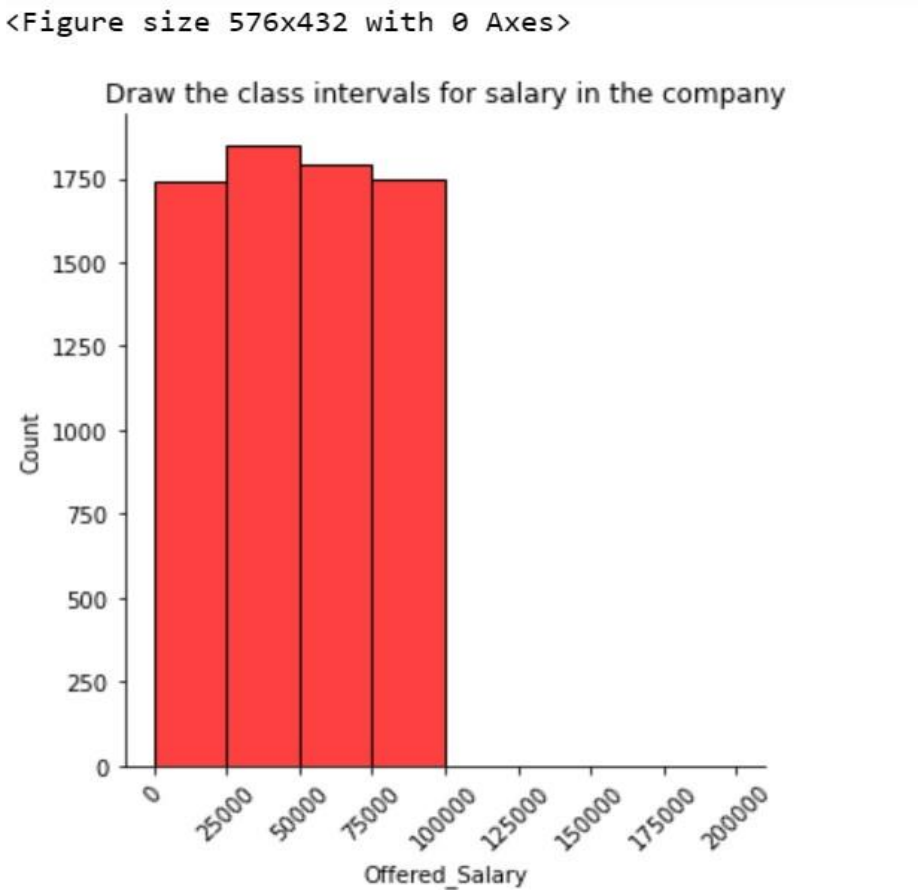
ans:- 85914

Post_Name	Average_Salary
-	85914.000000
b9	49847.287912
c-10	51244.359307
c5	50241.313003
c8	50747.257862
c9	50210.546884
i1	49937.954545
i4	48877.840909
i5	49467.559949

Class Intervals: The class interval is the difference between the upper class limit and the lower class limit.
Your task: Draw the class intervals for salary in the company ?

- code used-

```
plt.figure(figsize=(8,6))
sns.displot(data["Offered_Salary"],
bins=[0,25000,50000,75000,100000,125000,150000,175000,200000]
,color="red")
plt.title("Draw the class intervals for salary in the company")
locs,labels=plt.xticks()
plt.setp(labels, rotation=45)
plt.show()
```

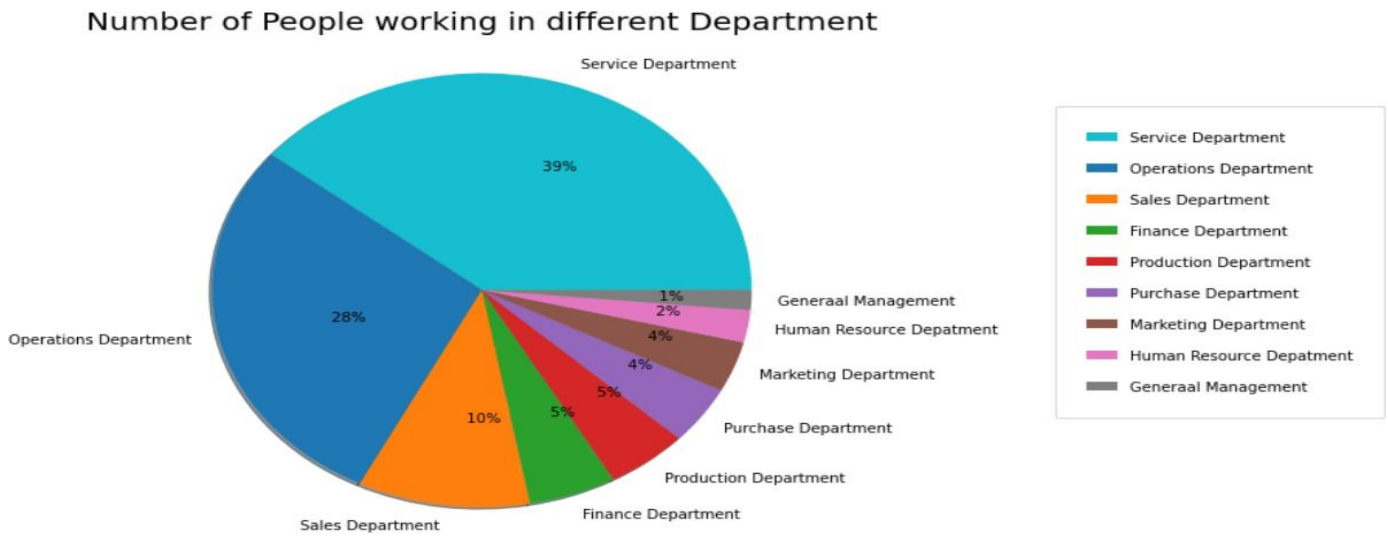


Charts and Plots: This is one of the most important part of analysis to visualize the data.

Your task: Draw Pie Chart / Bar Graph (or any other graph) to show proportion of people working different department ?

- Code used:-

```
Dep=["Service Department","Operations Department","Sales Department","Finance Department",
    "Production Department","Purchase Department","Marketing Department", "Human Resource Depatment", "Generaal Management"]
total=c.Department.value_counts() explode=(0,0,0,0,0,0,0,0,0) plt.pie(total) explode=(0,0,0,0,0,0,0,0,0)
plt.pie(total,explode=explode,labels=Dep,autopct='%12.0f%%',shadow=True,radius = 2,startangle=360)
plt.title('Number of People working in different Department',fontsize = 20,pad=100.0)
plt.legend(loc="upper right",handlelength=2, borderpad=2, labelspacing=1.5,bbox_to_anchor=(3.2,1.2))
plt.show()
```

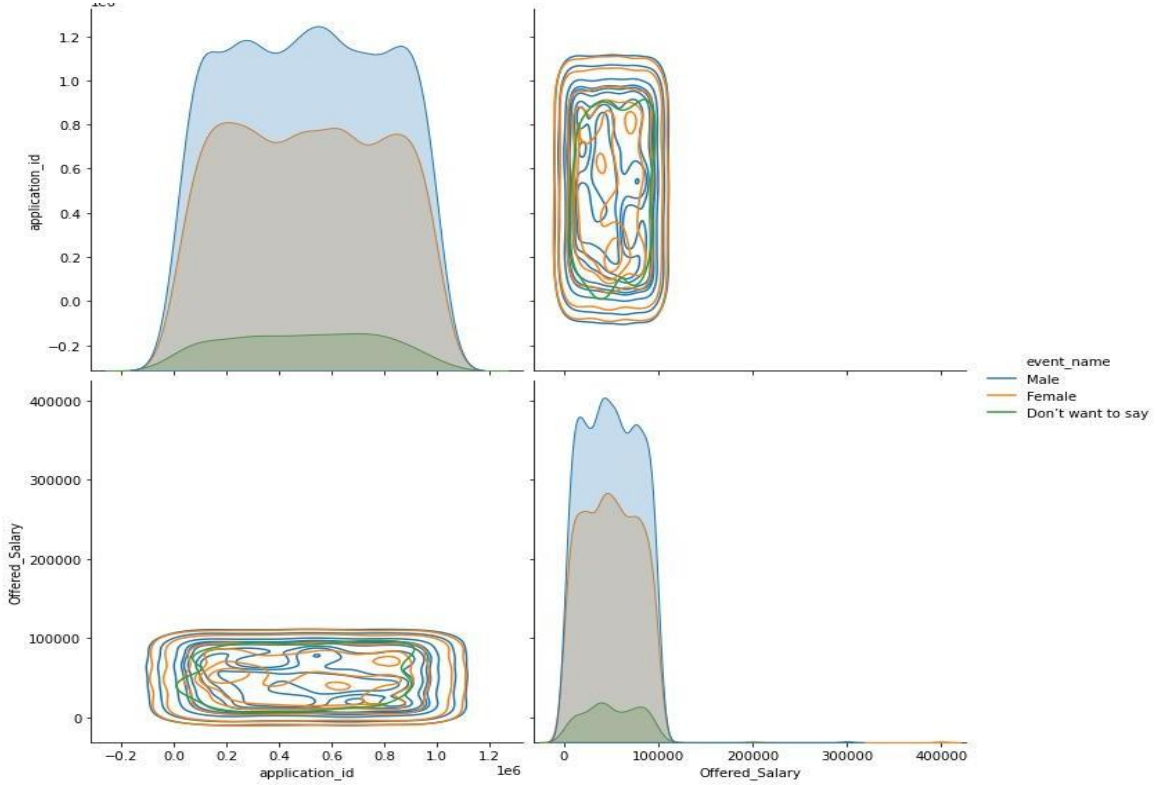


Charts: Use different charts and graphs to perform the task representing the data.

Your task: Represent different post tiers using chart/graph?

- Code used:-

```
sns.pairplot(c,hue="event_name“
,kind="kde",height=5)
plt.show()
```



Result-

- I got a chance to work with real time dataset. Also I got chance to again work with python tools and libraries. I learned more techniques in jupyter notebbok.
- I have answered all the questions asked in the data set and tried to plot the required graphs and chat as per requirement and my understanding. This project has helped in getting a hands-on experience on real life data set and how we clean, manipulate, visualize, and draw insights from the data.
- Exploratory Data Analysis part has helped me to understand that before moving towards making further analytical treatment of data and making it fit for making models. We have to do the EDA to make the data error free and bias free so that the inference drawn from the data is a good a fit to the further statistical and analytical treatment.
- I understood how to derive insights from a dataset that we are given.

➢ Analysis File : link

➢ Dataset : link