Assignment Report

1. Column Descriptions

- **name:** Names of individuals
- **age:** Age of individuals
- **joining_date:** Date of joining(likely)
- **income:** Monthly income
- **job_category**: Coded values (A, B, C) indicating job type
- **remote:** Binary indicators of job being remote
- **unknown1:** Unidentified field
- **city:** City names
- **country:** Country names
- **unknown2:** Unidentified field
- **gender:** Coded gender values (‘M’, ‘F’, ‘Other’)
- **phone:** Phone numbers
- **email:** Email addresses
- **marital_status:** Marital state
- **employment_status:** Employment state (e.g. employed, unemployed)
- **job_role:** Job titles (e.g. Contractor, Web Designer)
- **analytics:** Tools like Python, SQL
- **dashboard:** Tools like Tableau, PowerBI
- **experience:** Years of experience
- **company:** Company names

2. Assumptions

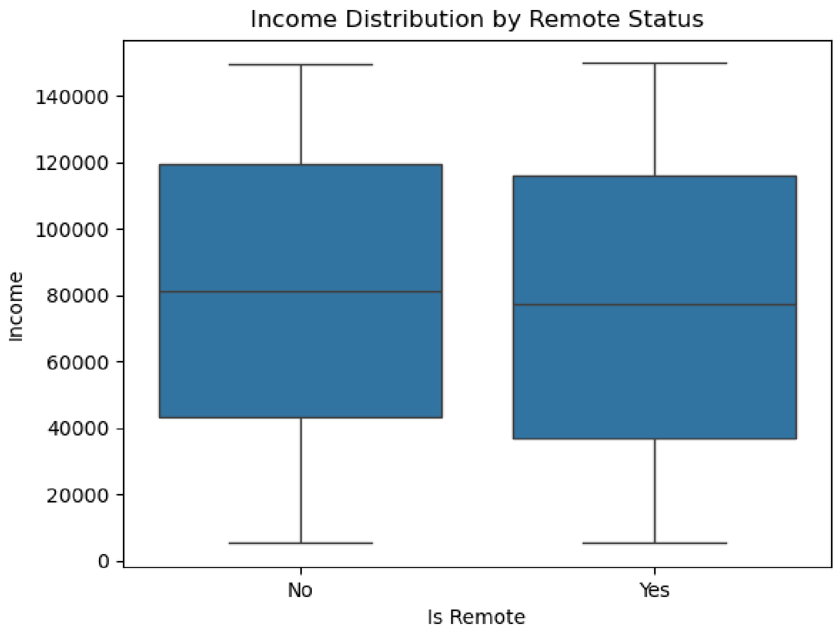
- The date column is assumed to be joining date, but it won’t make a difference in the analysis.
- Column containing entries like (a, b, c, A, B, C) is interpreted as job category.
- One binary column was wildly guessed to be remote column by considering the nature of its neighboring columns.
- Job-related columns may represent prior employment (or data will be contradictory).

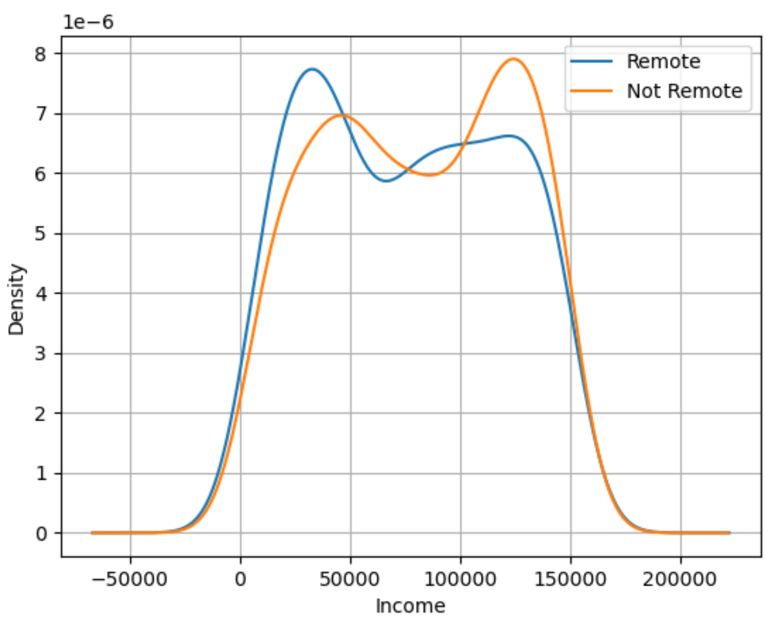
3. Cleaning Steps

- Converted age, experience, income to int32.
- Parsed joining_date to datetime format.
- Extracted monthly income from Json embedded text with simple str operation
- Standardized and converted job_category, remote, gender, employment_status, analytics, dashboard to categorical types.
- Fixed inconsistencies in values (e.g. 'Yes', 'yes', 'Y', etc.).

4. Key Insights

1. Income Distribution by Remote Status

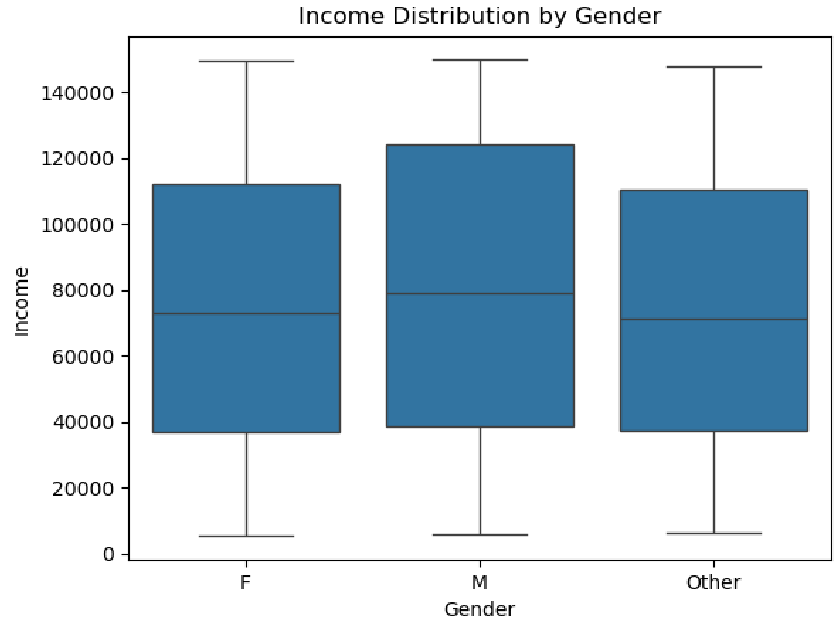


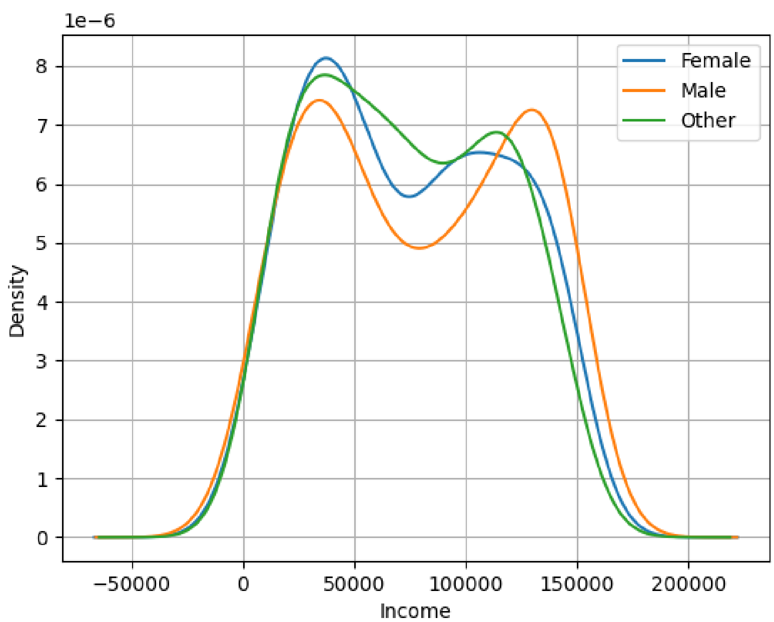


Insight:

- Avg remote salary – 76k, Avg non remote salary – 81k
- With a salary of more than 100k a job is more likely to be a non-remote job which can be because of companies wanting their employees to come to office when they are paid high, whereas for a job of less than 50k the company is more open to remote positions, which also explains the higher avg salary in non-remote jobs.

2. Income Distribution by Age

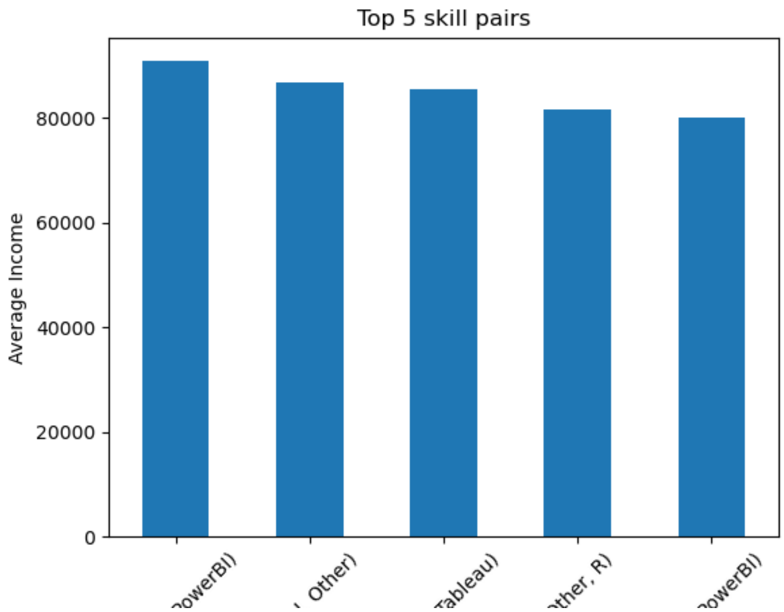




Insight:

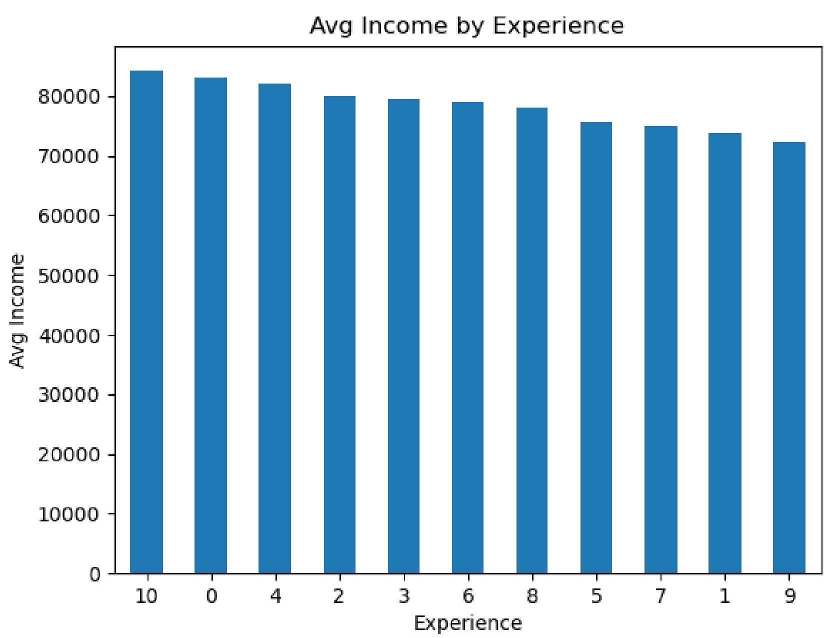
- On average a man earns 5.4% more than a woman and 7.6% more than other genders.
- Proportion of female and other genders are more than male in jobs with salary less than 120k, but men dominate in jobs with more than 120k salary explaining the avg. pay disparity.

3. Income Distribution by Skills



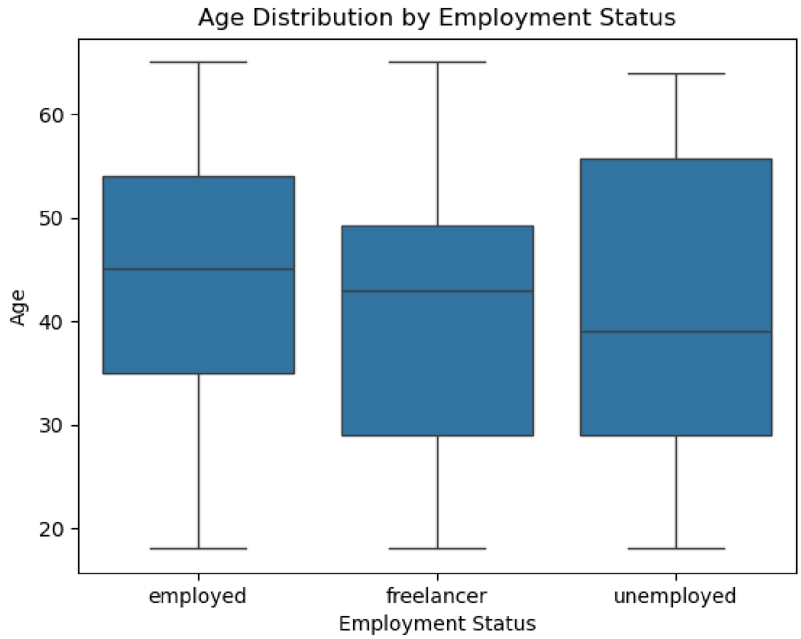
Insight: people who knows Python and Power BI together earns almost 14% more on an average than the average income.

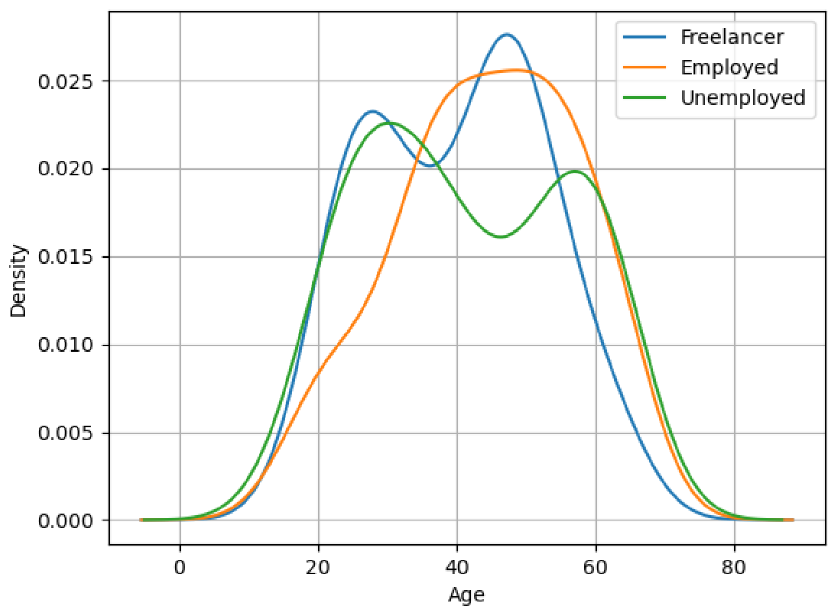
4. Income Distribution by Experience



Insight: It is often believed that more experience equals more income. Although the avg income highest for 10 years of experience, it is 2nd highest for 0 years of experience and lowest for 9 years of experience. So, clearly contrary to the popular belief the income is distributed randomly by experience and doesn't show any relation at all.

5. Age Distribution by Employment Status

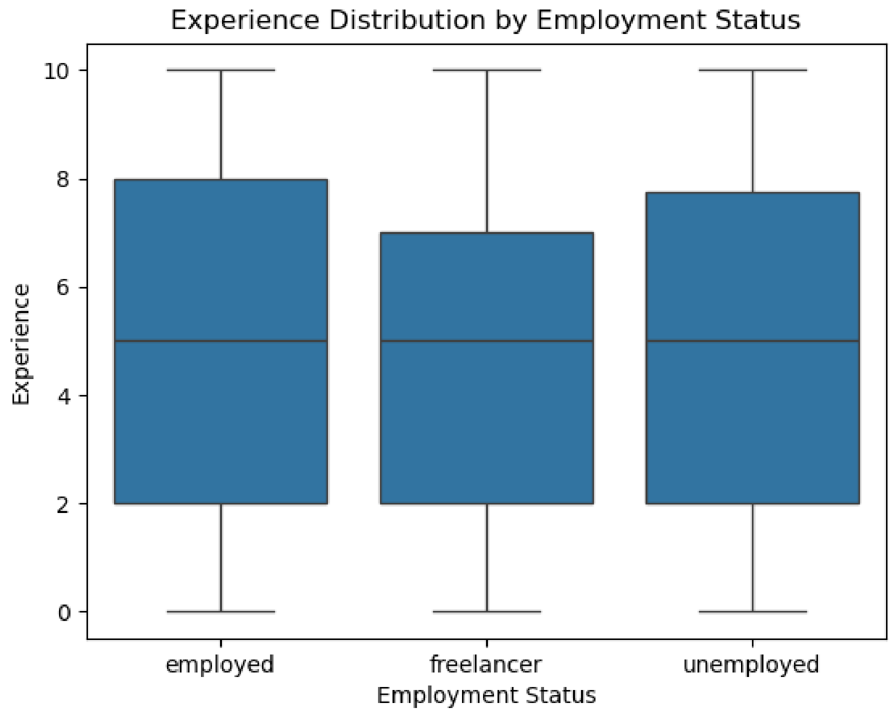


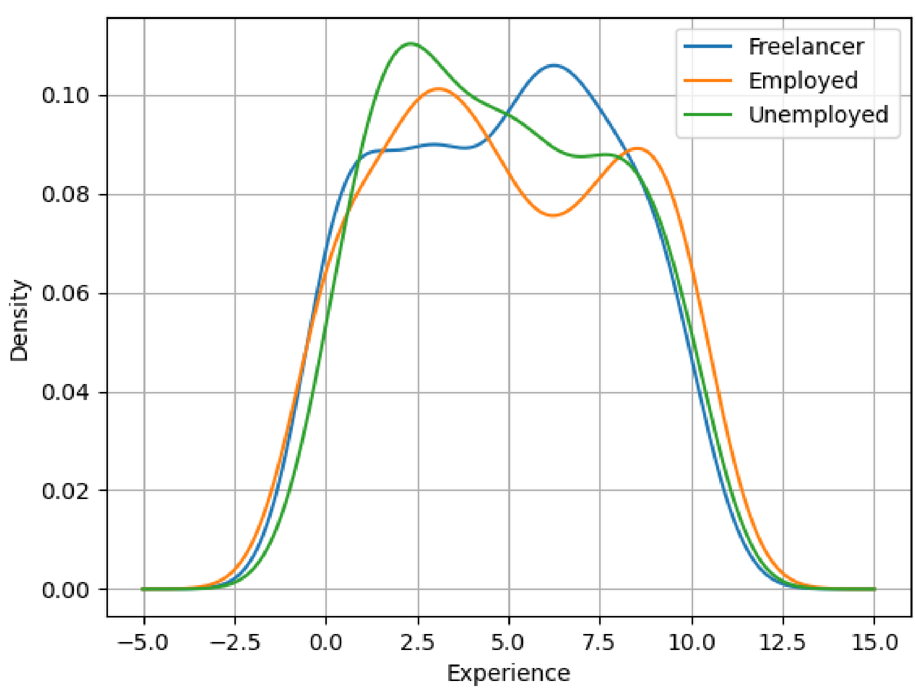


Insight:

- In the age group 18 to 35 people are more likely to be unemployed or working as a freelancer.
- In the age group 35 to 60 people are less likely to be unemployed which explains the avg. age of employed peoples being higher.

6. Experience Distribution by Employment Status

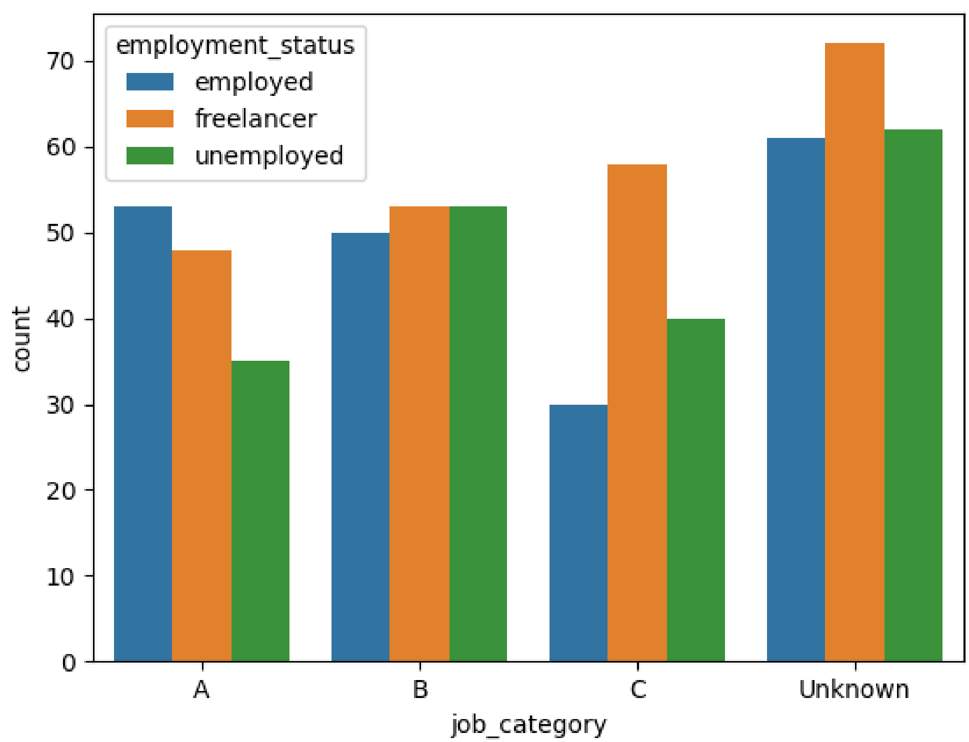




Insight: People with less than 5 years of experience are more likely to be unemployed whereas people with more than 5 years of experience people tends to shift towards freelancing.

7. Employment Status by Job Category

job_category	A	B	C	Unknown
employment_status				
employed	38.970588	32.051282	23.4375	31.282051
freelancer	35.294118	33.974359	45.3125	36.923077
unemployed	25.735294	33.974359	31.2500	31.794872



Insight: Only 25% of people who worked on a job with category A are unemployed currently, whereas in category B and C only almost 34% and 31% are unemployed. This can follow from the fact that group A has a lower income than other categories, meaning less layoffs from group A.

Dashboard

Streamlit dashboard link: Dashboard