Assignment 03

Gavin Boss

September 20, 2025

# 1. Companies Table

|  | company\_name | company\_raw | company\_is\_staffing | company\_id |
| --- | --- | --- | --- | --- |
| 0 | Crowe | Crowe | False | 0 |
| 1 | The Devereux Foundation | The Devereux Foundation | False | 1 |
| 2 | Elder Research | Elder Research | False | 2 |
| 3 | NTT DATA | NTT DATA Inc | False | 3 |
| 4 | Frederick National Laboratory For Cancer Research | Frederick National Laboratory for Cancer Research | False | 4 |

# 2. Data Preparation (Clean Up Data)

Medians : 87295.0 130042.0 115024.0  
Data cleaning complete. Rows retained: 72498

|  | EMPLOYMENT\_TYPE\_NAME | SALARY |
| --- | --- | --- |
| 0 | Part-time / full-time | 92500.0 |
| 1 | Full-time (> 32 hours) | 110155.0 |
| 2 | Full-time (> 32 hours) | 92962.0 |
| 3 | Full-time (> 32 hours) | 107645.0 |
| 4 | Full-time (> 32 hours) | 192800.0 |

|  | EMPLOYMENT\_TYPE\_NAME | SALARY |
| --- | --- | --- |
| 0 | Part-time / full-time | 92500.0 |
| 1 | Full-time (> 32 hours) | 110155.0 |
| 2 | Full-time (> 32 hours) | 92962.0 |
| 3 | Full-time (> 32 hours) | 107645.0 |
| 4 | Full-time (> 32 hours) | 192800.0 |

EMPLOYMENT\_TYPE\_NAME  
Full-time (> 32 hours) 116500.0  
Part-time ( 32 hours) 86390.0  
Part-time / full-time 100000.0  
Name: SALARY, dtype: float32

EMPLOYMENT\_TYPE\_NAME  
Full-time (> 32 hours) 116500.0  
Part-time ( 32 hours) 86390.0  
Part-time / full-time 100000.0  
Name: SALARY, dtype: float32

# 3. Salary Distribution by Industry and Employment Type

Compare salary variations across industries.  
Filter the dataset  
 Remove records where salary is missing or zero.  
Aggregate Data  
 Group by NAICS industry codes.  
 Group by employment type and compute salary distribution.  
Visualize results  
 Create a box plot where:  
 X-axis = NAICS2\_NAME  
 Y-axis = SALARY\_FROM  
 Group by EMPLOYMENT\_TYPE\_NAME.  
 Customize colors, fonts, and styles.  
Explanation: Write two sentences about what the graph reveals.

Two sentencences that the graph reveals - This graph reveals the high and low end of jobs while displaying the outliers in an easy to read graphic for specific industries. Noted here is the large difference in accommodation and food services, which can have a large range from 100k to as high as roughly 200k median, while some other jobs such as education services could have a median of roughly 60k to 80k.

# 4. 3 Salary Analysis by ONET Occupation Type (Bubble Chart)

–Appendix 1: Asked Copilot to help, as my aggregation was not workiong correctly, but it was because of a mix of the aggregation and the sorting that we had done in the saturday help session. AI prompts attached.

Analyze how salaries differ across ONET occupation types.  
Aggregate Data  
 Compute median salary for each occupation in the ONET taxonomy.  
Visualize results  
 Create a bubble chart where:  
 X-axis = ONET\_NAME  
 Y-axis = Median Salary  
 Size = Number of job postings  
 Apply custom colors and font styles.  
Explanation: Write two sentences about what the graph reveals.

Two sentencences that the graph reveals - This graph reveals that Data analysts may have the biggest field and the most data sets, but enterprise architexts make the most money on average based on the jobs in this data set.Speciality jobs that have specific skills also seem to have higher pay as opposed to general jobs like business analyst, unlike business intelligence analysts who have a higher average salary.

# 5. 4 Salary by Education Level (Two Groups)

Create two groups:  
 Associate’s or lower (GED, Associate, No Education Listed)  
 Bachelor’s (Bachelor’s degree)  
 Master’s (Master’s degree)  
 PhD (PhD, Doctorate, professional degree)  
Plot scatter plots for each group using, MAX\_YEARS\_EXPERIENCE (with jitter), Average\_Salary, LOT\_V6\_SPECIALIZED\_OCCUPATION\_NAME  
After each graph, add a short explanation of key insights.

Two sentencences that the graph reveals - This graph reveals that masters programs can aid in getting you higher paying jobs in the beginning as an aid in replacing years of experience, and even over the longer term has shown in this graph to translate to higher average salaries. This could be multiple factors such as company requirements, but there are also outliers, such as the masters degree with 10 years of experience that has an average salary of roughly 60k.

# 6. 4 Salary by Education Level (Four Groups)

Create two groups:  
 Associate’s or lower (GED, Associate, No Education Listed)  
 Bachelor’s (Bachelor’s degree)  
 Master’s (Master’s degree)  
 PhD (PhD, Doctorate, professional degree)  
Plot scatter plots for each group using, MAX\_YEARS\_EXPERIENCE (with jitter), Average\_Salary, LOT\_V6\_SPECIALIZED\_OCCUPATION\_NAME  
After each graph, add a short explanation of key insights.

See appendix 2 – asked ai to help me fix the data being in a straight line and it suggested the jitter.

Two sentencences that the graph reveals - This graph indicates that early in careers PhD holders can have an increased salary in the few cases noted, but over the long term years of experience translates to higher pay. Experience takes over as a key factor to salary according to this data with higher years of experience focusing on associates or lower and bachelors as opposed to higher level degrees.

|  | MAX\_YEARS\_EXPERIENCE | Average\_Salary | LOT\_V6\_SPECIALIZED\_OCCUPATION\_NAME | REMOTE\_GROUP |
| --- | --- | --- | --- | --- |
| 0 | 2.0 | 92962.0 | Data Analyst | Onsite |
| 1 | 2.0 | 75026.0 | Oracle Consultant / Analyst | Onsite |
| 2 | 1.0 | 60923.0 | Data Analyst | Remote |
| 3 | 2.0 | 131100.0 | Enterprise Architect | Onsite |
| 4 | 3.0 | 136950.0 | Enterprise Architect | Remote |

Two sentencences that the graph reveals - This graph reveals that there isn’t necessarily a direct correlation between years or experience, or a difference in salary based on remote work vs onsite work. There are higher salaries around the 7 year mark, but there is roughly an even looking mix based on the data in groups that indicate relatively equal pay between onsite and remote workers, with hybrid workers remaining in that median as well.

# 7. Histrograms

# 8. Reference Appendix 3 for AI Prompt, had data but needed help formulating code to get all histographs in one.

Remote Jobs- For remote jobs, there seem to be less lower paying jobs, to me this means that the pattern is these are higher skilled jobs or jobs that require more experience, and allow for more flexibility in work location, which is why we see a gap in the left side of the histogram and it starts to make a mountain towards the 100k salary range. The same can be true though for the high paying positions as there are very few of them as well.

Hybrid Jobs - A similar trend here is seen only is is shifting to the right, where there are more jobs that are paying a but less and the peak is at around the 80k range as opposed to the remote job. There are also less data points to gather this information from as there are only a max of roughly 25 jobs in a maximum salary range for the hybrid jobs as opposed to remote and onsite jobs.

Onsite Jobs - Onsite jobs have the largest share of the job market and the salary ranges are fairly even, they are tapered off towards the 200k mark but there isn’t a large spike in the number of job count like there were seen in hybrid or remote jobs, but there is also a lot more data to look at when looking at onsite jobs.