Assignment 03

Dominique Strong

September 24, 2025

|  |
| --- |
| Data Importing and Preparation |
| :: {.cell execution\_count=2} |
| :: {.cell-output .cell-output-stderr} ARNING: Using incubator modules: jdk.incubator.vector sing Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties etting default log level to "WARN". o adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel). 5/09/24 03:14:15 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable 5/09/24 03:14:29 WARN SparkStringUtils: Truncated the string representation of a plan since it was too large. This behavior can be adjusted by setting 'spark.sql.debug.maxToStringFields'. :: :: |
| :: {.cell execution\_count=6} |
| :: {.cell-output .cell-output-stderr} Stage 4:> (0 + 1) / 1] :: |
| :: {.cell-output .cell-output-stdout} edians: SALARY\_FROM = 87295.0, SALARY\_TO = 130042.0, SALARY = 115024.0 :: |
| :: {.cell-output .cell-output-stderr} `` |
| `` :: :: |
| :: {.cell execution\_count=9} |
| :: {.cell-output .cell-output-stderr} `` |
| `` :: :: |
| :: {.cell execution\_count=10} |
| :: {.cell-output .cell-output-stderr} `` |
| `` :: :: |

#1) Salary Distribution by Industry and Employment Type

Unable to display output for mime type(s): text/html

Unable to display output for mime type(s): text/html

This graph illustrates how salaries vary by each NAIC, where Administrative, Support, Waste Management and Remediation services has the highest salaries compared to all other NAIC’s at $500K. Conversely, Unclassified has the lowest salary available at roughly $15K.

While many NAIC’s have longer ranges of salaries, such as Information, Public Admin, and Real Estate, NAIC’s such as Retail Trade mostly consist of the median salary of $115K, therefore any salaries above or below that are considered outliers.

#2) Salary Analysis by ONET Occupation Type (Bubble Chart)

|  | Occupation\_Name | Median\_Salary | Job\_Postings |
| --- | --- | --- | --- |
| 0 | Data / Data Mining Analyst | 95250.0 | 30057 |
| 1 | Business Intelligence Analyst | 125900.0 | 29445 |
| 2 | Computer Systems Engineer / Architect | 157600.0 | 8212 |
| 3 | Business / Management Analyst | 93650.0 | 4326 |
| 4 | Clinical Analyst / Clinical Documentation and ... | 89440.0 | 261 |

Unable to display output for mime type(s): text/html

This graph shows that Data Mining and Business Intelligence analysts have the highest number of job postings, where the former has a lower median salary compared to the former. Computer Systems Engineer however, has a moderate number of job postings at a higher median salary, whereas Business, Clinical, and Market Research Analyst have the lowest number of job postings at the lowest salaries.

#3) Salary by Education Level

Unable to display output for mime type(s): text/html

This graph illustrates how salary varies by education level and experience - after close examination, it is clear that as the years of experience increases, so does the average salary. For example, for all educations levels, when their is more than 3 years of experience, the average salary tends to fall somewhere between $100K and $200K; conversely 1 year of experience average salaries tend to fall more below $100K. Interestingly, a Bachelor degree seems to dominate more salaries in the 6+ years of experience.

#4) Salary by Remote Work Type

Unable to display output for mime type(s): text/html

Similar to the prior graph, this graph compares average salaries across years of experience but now with the type of work situation: onsite, remote, or hybrid. One key callout, the highest average salaries for most (if not all) years of experience is when the work is onsite - onsite workers tend to get paid higher on average compared to remote, whereas hybrid salaries fluctuate the most.