DATA SCIENTIST

GREG KHOMAR 2030

WHAT THEY DO

DATA SCIENTISTS APPLY ADVANCED STATISTICAL AND COMPUTATIONAL METHODS TO EXTRACT KNOWLEDGE AND INSIGHTS FROM STRUCTURED AND UNSTRUCTURED DATA.

- **CLEAN AND ORGANIZE DATA:** SPEND SIGNIFICANT TIME COLLECTING, PROCESSING, AND "CLEANING" RAW DATA (OFTEN 60-80% OF A PROJECT) TO ENSURE IT'S ACCURATE AND READY FOR ANALYSIS.
- **DEVELOP MODELS & ALGORITHMS:** BUILD, TEST, AND VALIDATE **MACHINE LEARNING (ML)** MODELS AND PREDICTIVE ALGORITHMS TO FORECAST TRENDS, CLASSIFY DATA, OR OPTIMIZE BUSINESS PROCESSES.
- **COMMUNICATE FINDINGS:** CREATE AND PRESENT COMPELLING DATA VISUALIZATIONS, REPORTS, AND STRATEGIC RECOMMENDATIONS TO TECHNICAL AND NON-TECHNICAL STAKEHOLDERS TO DRIVE DATA-DRIVEN DECISION-MAKING.

ED UCATION & SKILLS

Category	Requirement/Skills
Minimum Education	Bachelor's degree (Computer Science, Statistics, Mathematics, Data Science, or a related quantitative field).
Recommended Education	Master's degree in Data Science, Computer Science, or a closely related field is highly valued and often preferred for specialized or senior roles.
Technical Skills	1. **Programming Proficiency:** Mastery of **Python** (for ML, libraries like Pandas, NumPy, Scikit-learn) and **R** (for statistical analysis). 2. **Database Querying:** Strong knowledge of **SQL** for extracting and manipulating data from relational databases. 3. **Machine Learning/AI:** Deep understanding of ML algorithms (regression, classification, clustering, deep learning) and model deployment.
Soft Skills	1. **Problem-Solving/Analytical:** The ability to devise creative, data-backed solutions to complex, ambiguous business problems. 2. **Communication & Storytelling:** Skill in translating complex analytical results into clear, actionable insights for non-technical audiences. 3. **Curiosity & Continuous Learning:** A strong desire to explore new data, tools, and algorithms in a rapidly evolving field.

Salary snapshot

Career Level	Average Base Salary
Entry-Level (0-2 yrs)	$\approx \$100,369$
Mid-Level (2-5 yrs)	$\approx \$134,182$
Senior-Level (6+ yrs)	$\approx \$146,430$

(Source: Interview Query/Glassdoor/BLS, based on various salary reports)

JOB OUTLOOK

- **Growth Rate: 33.5%** projected growth from 2024 to 2034, significantly **much faster than the average** for all occupations (U.S. Bureau of Labor Statistics).

- **Demand Summary:** Demand is exceptionally strong and is driven by the continued exponential growth of data (**Big Data**) and the increasing need for organizations across all sectors (tech, finance, healthcare, etc.) to use **AI** and data analysis to make informed decisions, develop new products, and gain a competitive edge.

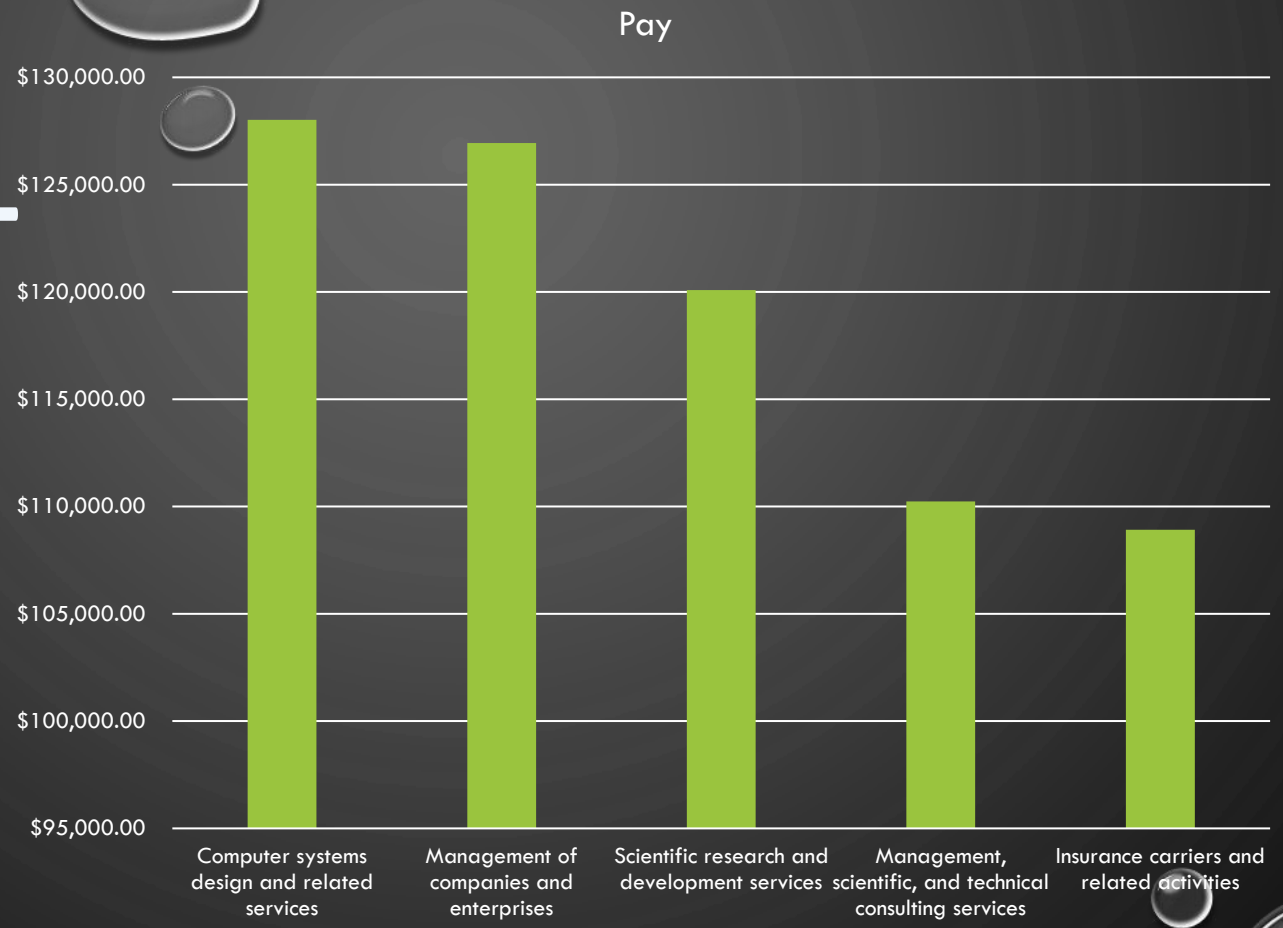
COURSEWORK CONNECTION

Code: **Essential-**
Programming languages like Python and R are the **primary tools** for data cleaning, statistical analysis, model development, and scripting data pipelines.

DB (Database)Critical-
Knowledge of **SQL** and database concepts (data warehousing, data modeling) is necessary to efficiently query, retrieve, and structure the large datasets used for analysis.

AI (Artificial Intelligence)Core-
The entire field relies on Machine Learning (ML)—a major subset of AI—to build predictive and prescriptive models. Courses in Deep Learning, NLP (Natural Language Processing), and MLOps are directly applicable.

TREND CHART



Data Scientists : Occupational Outlook Handbook: : U.S. Bureau of Labor Statistics

DAY-IN-THE-LIFE

DATA PREP: SPEND SIGNIFICANT TIME WRITING **PYTHON** OR **SQL** SCRIPTS FOR DATA CLEANING AND TRANSFORMATION (**DATA WRANGLING**).

MODELING: BUILD, TRAIN, AND TEST **MACHINE LEARNING** MODELS AND ALGORITHMS, FOLLOWED BY DEBUGGING AND ITERATION.

COMMUNICATION: PARTICIPATE IN BRIEF DAILY **STANDUPS** (AGILE) AND CONDUCT **STAKEHOLDER REVIEWS** TO PRESENT ANALYSIS, VISUALIZATIONS, AND FINDINGS TO BUSINESS LEADERS.

Category	**Examples**
Programming	**Python** (Pandas, Scikit-learn), **R**
Databases	**SQL**, Snowflake, Redshift
Cloud	AWS, GCP, Microsoft Azure
Visualization	Tableau, Power BI, Matplotlib
Development	Jupyter Notebooks, Git/GitHub

REFERENCES

- U.S. BUREAU OF LABOR STATISTICS (BLS)
- INTERVIEW QUERY SALARY REPORT
- VARIOUS ACADEMIC/INDUSTRY REPORTS