

Mapping the Data Science Landscape

Chiranthan Shadaksharaswamy

INTRODUCTION

Mapping the Data Science Landscape, the burgeoning field of data science is rapidly transforming businesses across the globe. With growing demand for skilled professionals, understanding the complexities of the data science job market is more crucial than ever. This project delves into the intricate web of data science roles, salaries, and global trends, leveraging comprehensive data analysis to provide valuable insights for both job seekers and stakeholders.

	A	B	C	D	E	F	G	H	I	J	K	L
1	work_year	job_title	job_category	salary_currency	salary	salary_currency	employee_experience	employment_type	work_setting	company_location	company_size	company_type
2	2023	Data Dev Data Engr	EUR	88000	95012	Germany	Mid-level	Full-time	Hybrid	German	L	
3	2023	Data Arc Data Arc	USD	2E+05	2E+05	United Sta	Senior	Full-time	In-persc	United	5	M
4	2023	Data Arc Data Arc	USD	81800	81800	United Sta	Senior	Full-time	In-persc	United	5	M
5	2023	Data Scie Data Scie	USD	2E+05	2E+05	United Sta	Senior	Full-time	In-persc	United	5	M
6	2023	Data Scie Data Scie	USD	93300	93300	United Sta	Senior	Full-time	In-persc	United	5	M
7	2023	Data Scie Data Scie	USD	1E+05	1E+05	United Sta	Senior	Full-time	Remote	United	5	M
8	2023	Data Scie Data Scie	USD	1E+05	1E+05	United Sta	Senior	Full-time	Remote	United	5	M
9	2023	Machine Machine	USD	2E+05	2E+05	United Sta	Mid-level	Full-time	In-persc	United	5	M
10	2023	Machine Machine	USD	1E+05	1E+05	United Sta	Mid-level	Full-time	In-persc	United	5	M
11	2023	Data Eng Data Engr	USD	2E+05	2E+05	United Sta	Executive	Full-time	Remote	United	5	M
12	2023	Data Eng Data Engr	USD	2E+05	2E+05	United Sta	Executive	Full-time	Remote	United	5	M
13	2023	Machine Machine	USD	2E+05	2E+05	United Sta	Senior	Full-time	In-persc	United	5	M
14	2023	Machine Machine	USD	1E+05	1E+05	United Sta	Senior	Full-time	In-persc	United	5	M
15	2023	Data Scie Data Scie	GBP	35000	43064	United Kin	Mid-level	Full-time	In-persc	United	5	M
16	2023	Data Scie Data Scie	GBP	30000	36912	United Kin	Mid-level	Full-time	In-persc	United	5	M
17	2023	Data Ana Data Ani	USD	95000	95000	United Sta	Entry-level	Full-time	In-persc	United	5	M
18	2023	Data Ana Data Ani	USD	75000	75000	United Sta	Entry-level	Full-time	In-persc	United	5	M
19	2023	Data Scie Data Scie	USD	3E+05	3E+05	United Sta	Senior	Full-time	In-persc	United	5	M
20	2023	Data Scie Data Scie	USD	2E+05	2E+05	United Sta	Senior	Full-time	In-persc	United	5	M

This project utilizes a comprehensive dataset encompassing job titles, salaries, experience levels, location, and company size, allowing for multi-faceted analysis. A combination of descriptive statistics, visualization techniques, and potential network analysis will be employed to uncover hidden patterns and trends. By shedding light on these critical aspects of the data science job market, this project hopes to empower individuals and organizations to make informed decisions in this ever-evolving field.

1 INSIGHT NEEDS

Our primary objective is to unveil the overarching trends within the dynamic field of data science, analysing the dataset through a temporal lens. By scrutinizing the work year column, we aim to extract valuable insights into the evolution of data-related job roles and their corresponding salaries over the years. This temporal understanding will empower stakeholders, allowing them to discern patterns, anticipate market shifts, and strategically position themselves within the ever-evolving data science landscape. As data science continues to mature, this historical context becomes a strategic asset, enabling stakeholders to make informed decisions regarding skill acquisition, workforce planning, and talent retention strategies. Additionally, understanding the temporal trajectory allows organizations to adapt to emerging technologies and industry trends, ensuring they stay at the forefront of innovation and maintain a competitive edge in the data science job market. Through this comprehensive temporal analysis, we not only aim to provide a snapshot of the current state of data-related job roles but also to offer a roadmap for stakeholders to navigate the future terrain of the data science landscape with confidence and foresight.

1.1 Stakeholder Analysis

Understanding the nuances of temporal trends in data-related job roles and salaries holds immense value for various stakeholders. HR professionals, tasked with talent acquisition and workforce planning,

will gain insights into the changing demands and market dynamics. Finance departments, concerned with budgeting and resource allocation, will find strategic guidance in anticipating salary trends. Data science leads, aiming to align roles with industry demands, can use this analysis to fine-tune their team structures. Executives, responsible for strategic decision-making, will benefit from a comprehensive view of the temporal evolution, aiding in long-term planning and resource allocation. By catering to the diverse needs of these stakeholders, our analysis aims to be a valuable compass guiding strategic decisions across different facets of the organization.

2 DATA ACQUISITION

Our journey into unravelling the intricacies of the data science landscape commences with a meticulous process of data acquisition. The dataset, thoughtfully sourced from Kaggle and organized in CSV format, stands as a reservoir of valuable insights. This comprehensive collection of information serves as the bedrock for our analysis, guiding us through the diverse facets of the data industry. Each row within the dataset represents a unique data point, offering a nuanced perspective on the ever-evolving dynamics of data-related roles, salaries, and employment conditions.

2.1 Description of Data

At the heart of the exploration lies a dataset rich in diversity and depth. In this tabular treasure trove, crucial data fields come to life, shaping our understanding of the data science job market. The "work_year" column introduces the temporal context, a cornerstone for unraveling salary trends over time. Job-specific details, encapsulated in the "job_title" and "job_category" fields, unlock insights into role-specific nuances and broader categorizations. Currency considerations are addressed through "salary_currency" and "salary" fields, offering a global perspective on compensation. Geographical nuances emerge from "employee_residence" and "company_location," influencing salary differentials and cost-of-living assessments. Professional dimensions, including experience levels, employment types, and work settings, are meticulously captured, providing a holistic view of employment dynamics. The dataset's structure, along with its varied fields, lays the groundwork for a comprehensive analysis, promising a wealth of insights into the multifaceted realm of data science careers.

3 ANALYSIS METHODS

As we navigate the intricate landscape of data science, our approach to analysis is multifaceted, drawing from a diverse array of methods to unveil the underlying patterns and insights within the dataset. The overarching goal is to provide stakeholders with a comprehensive understanding of temporal, geographical, and hierarchical dynamics within the data science job market. Our analysis methods are designed not only to illuminate current trends but also to project trajectories, empowering strategic decision-making in an ever-evolving field.

3.1 Temporal Analysis

The cornerstone of our approach lies in temporal analysis, leveraging the "work_year" column to dissect how data-related job roles and salaries have evolved over time. Through trend analysis and visualization techniques, we aim to unearth patterns, identify shifts in demand, and provide stakeholders with a historical context crucial for strategic decision-making.

3.2 Geospatial Analysis

Geospatial insights are derived from the "employee_residence," "company_location," and "salary_currency" fields. Mapping salary distributions, job prevalence, and regional trends allows us to uncover geographical patterns within the data science job market. This analysis

aids in understanding how salaries vary across countries and the impact of company location on compensation structures.

3.3 Topical Analysis

Our approach extends beyond traditional hierarchical structures, encompassing a topical analysis that delves into the "job_title," "job_category," and "company_size" fields. This methodology transcends organizational hierarchies, aiming to unravel topical intricacies within data-related roles. By leveraging tree or network analysis methods, we aim to unearth nuanced insights into role-specific dynamics, salary distributions, and the impact of company size on compensation. This topical analysis provides stakeholders with a comprehensive understanding of the intricate facets and specialized domains within the data science job market.

3.4 Tree and Network Analysis

In our analytical toolkit, we forego traditional correlation and regression analyses and instead opt for tree and network analysis methods. This innovative approach allows us to delve into the intricate relationships within the data science job market. By exploring connections between factors such as experience levels, job titles, and other categorical variables, our analysis aims to unveil complex hierarchical and networked structures that influence compensation dynamics. This method provides stakeholders with a unique perspective, shedding light on the interplay of factors shaping the nuanced landscape of salaries within the data science industry.

4 VISUALIZATIONS

Our commitment to delivering insightful and comprehensible findings extends to the realm of visualizations. These visual representations serve as powerful tools to convey complex trends, patterns, and relationships within the data science job market. The following subsections outline the visualizations and their prototypes that will enhance the interpretability of our analysis.

4.1.1 Temporal Analysis

Our exploration of temporal trends, anchored by the "work_year" column, unveils a rich tapestry of salary evolution within the data science job market. This temporal lens intertwines seamlessly with job-related details, such as job categories and company sizes, facilitating a nuanced analysis of how temporal trends interplay with other factors influencing salaries. Initial findings showcase a consistent upward trajectory in the average salaries of Data Analysts and Data Engineers, signalling growth trends. In contrast, roles associated with Machine Learning and AI exhibit more pronounced fluctuations, suggesting potential shifts in demand or market dynamics. The accompanying 100% stacked column chart visually delineates salary distribution across different roles in the broader field of Data Science, offering stakeholders a comprehensive snapshot of compensation dynamics. This visualization serves as a robust foundation for deeper analyses, empowering stakeholders to navigate the evolving landscape of data-related roles with informed decision-making.

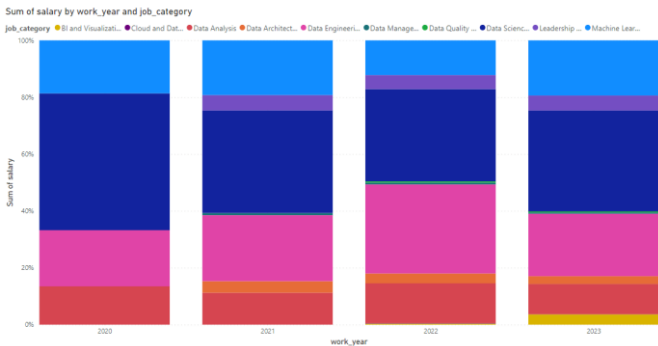


Fig. 1. Figure.

4.1.2 Geospatial Analysis

In our geospatial analysis, we delve into the complex terrain of salary distributions within the data science job market. Our chosen focus for visualization centres on mapping the distribution of job titles globally—a geospatial analysis providing insights into the geographic dispersion of various roles within the data science field. Initial findings reveal the United States as a hub of diversity, showcasing the highest variety in distinct data-related roles. The United Kingdom and Canada follow suit, demonstrating significant variability in the types of data roles present. These visualizations, crafted with Power BI, convey the diversity of data-related job titles across the global landscape, directly addressing stakeholder needs and paving the way for further exploration into regional salary trends and the intricate relationships between geographic locations and compensation dynamics.



Fig. 2. Figure.

5 INTERPRETATION OF RESULTS

The culmination of insights gathered over the four weeks unveils a nuanced narrative within the data science job market. Our temporal analysis portrays an upward trajectory in the salaries of Data Analysts and Data Engineers, indicative of robust growth trends. Conversely, the fluctuating patterns in salaries tied to Machine Learning and AI roles suggest dynamic shifts in demand or market dynamics. Geographical analyses spotlight the United States as a hub of diversity in data-related roles, closely followed by the United Kingdom and Canada. Hierarchical exploration deepens our understanding of organizational dynamics within data roles, while innovative tree and network analyses offer a unique perspective on the complex structures shaping compensation. Together, these findings create a comprehensive panorama of salary distributions, global job title diversity, and the impact of various factors on compensation levels. This interpretation serves as a strategic guide, empowering stakeholders to navigate the evolving data science landscape with informed decisions and proactive measures.

ACKNOWLEDGMENTS

The authors extend their gratitude to all classmates for their valuable insights during discussions, which played a pivotal role in enhancing and refining this project.

REFERENCES

[1] <https://www.kaggle.com/datasets/hummaamqaasim/jobs-in-data?resource=download>
[2] <https://learn.microsoft.com/en-us/power-bi/developer/visuals/>
[3] <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/temporal-analysis>
[4] <https://campus.datacamp.com/courses/introduction-to-power-bi/visualizing-data-3?ex=1>
[5] <https://cns.iu.edu/docs/presentations/2011-bomer-netsci-guestlecture.pdf>
[6] <https://www.storytellingwithdata.com/>