JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Navigating Data Science: Bridging Hype & Industry Realities

Manas Negi UG
Student (Data Science Department) Thakur College of Science & Commerce Kandivali (East), Mumbai

Saurav Yadav UG
Student (Data Science Department) Thakur College of Science & Commerce Kandivali (East), Mumbai

Mentor:
Dr. Gulab Nibrad,
Department of Statistics,
Thakur College of Science & Commerce Kandivali (East), Mumbai

Rohan Gupta UG
Student (Data Science Department) Thakur College of Science & Commerce Kandivali (East), Mumbai

Sapna Samanta UG
Student (Data Science Department) Thakur College of Science & Commerce Kandivali (East), Mumbai

I. Abstract: -

In recent years, the field of Data Science has surged in popularity, captivating the interest and aspirations of both students and professionals across various industries. The allure of Data Science, driven by its perceived significance in decision-making and transformative potential, has propelled it into the spotlight as a lucrative and impactful career choice. However, amidst this fervour, a critical gap seems to emerge between the enthusiastic hype surrounding Data Science and the pragmatic realities of its demand within the job market. This research embarks on an exploratory journey to unravel the sophisticated dynamics between the inflated perceptions of Data Science careers and the actual industry demand. By engaging with two pivotal stakeholders – students contemplating Data Science as a career path or amid transitioning, and professionals who have already entrenched themselves within the field – this study aims to extract valuable insights into their expectations, misconceptions, and informed perspectives. Employing a meticulously structured approach, encompassing surveys and data collection, this research delves into the core factors that contribute to the heightened buzz around Data Science. It seeks to decipher whether the allurement of Data Science as a prosperous and impactful occupation authentically mirrors the noticeable conditions of its demand and prospects. The comparison between the anticipated allure and real-world requisites of Data Science careers forms the crux of this investigation. Through comprehensive data analysis and interpretation, this study endeavors to bridge the gap between perception and reality. By discerning the underlying reasons for the prevailing hype, this research intends to equip aspiring students and industry professionals with a well-rounded understanding, enabling them to make informed decisions aligned with the actual landscape of Data Science careers. Furthermore, the outcomes of this study could potentially contribute to

refining educational curricula, industry expectations, and individual career strategies in the context of Data Science

Keywords - Disparity, Expectations, Misconceptions, Data Analysis, Informed Perspectives, Heightened Buzz, Perceptions and Reality.

II. Introduction –

With the rapid development in the IT sector the amount of data namely the Data Volume keeps on increasing and so does the need for personnel who can understand the data and to create inferences out of it. According to a study [amtdata] (Duarte), the amount of data generated in the year 2023 will be around 120 zettabytes and this amount will keep on increasing each year.

Hence, Data Science became a much-needed requirement in the industry. Seeing the rising popularity of Data Science in the industry, educational platforms started providing Data Science as an academic course.

We became aware of this widespread popularity of Data Science and its growing significance. Intrigued by the buzz surrounding this field, we decided to delve deeper to ascertain the validity of its popularity and explore the reasons behind its rising prominence.

Prof. Jeff Wu from the University of Michigan called for statistics to be renamed Data Science and statisticians to be called Data Scientists. Data Science is not entirely a new field, it is not a single discipline it contains the aspects of multiple disciplines like statistics and computer science.

To the best of our knowledge, we didn't find any research paper explaining the hype of data science and the industry realities in this field. Therefore, we decided to perform research on the exaggerated hype about data science and the actual reality of it in the industry.

III. Methodology –

The current study is based on data collected from various sources such as survey forms, interviews, websites, etc. There were two sources for our data collection primarily consisting of Primary data collection and secondary data collection methods. Our survey consisted of two main components, the first being the students who are currently pursuing data science and the students who are interested in data science as a future career path choice. The second part consists of data science professionals who are currently in the data science domain and any other professionals who are in the IT sector. For interviews, we tried reaching out to professionals who are currently working in the Data Science domain through social media platforms like LinkedIn and requesting them to have a Q&A session with us.

IV. Literature Survey

The student pursuing data science as a career path and the professionals who are trying to transition into the data science domain can pursue various job roles such as:

Data Scientist:

A Data Scientist explores various data patterns to measure the impact on an organization. A key role of a Data Scientist is the ability to explain the importance of data in a simpler method to be understood by others. They are supposed to have a statistical knowledge of different programming languages required for solving complex problems. (Sharma)

Data Analyst:

Analyzing data to figure out a market trend is the role of a Data Analyst. He helps in providing a clear picture of the company's standing in the market. Once the desired goal is set by a company, a Data Analyst provides datasets to achieve the required goal. (Sharma)

Data Engineer:

The Data Engineer works with the core of the organization and can be considered the backbone of a company. They are the builder, designers, and managers of a large database. They are in charge

of building data pipelines, enabling correct data flow, and ensuring the data reaches the relevant departments. (Sharma)

Business Intelligence Analyst:

A business intelligence analyst helps in analyzing the collected data to maximize the company's efficiency, hence generating more profits. Their role is more technical than analytical, requiring more knowledge of popular machines. They have to serve as a bridge between business and IT, helping them improve. (Sharma)

Marketing Analyst:

The role of a marketing analyst is to assist companies in their marketing division. They analyze and suggest which product to produce in large quantities and which product to discontinue. Monitoring customer satisfaction reports helps in improving existing products and services. They decide which products to sell to the targeted customers and at which price. (Sharma)

Data Architect:

One of the most essential careers in data science is to become a data architect. A data architect designs, develops and maintains a business data management system. They are responsible for meeting the company's database requirements and building them in keeping with internal and external regulations. (Sharma)

Machine Learning Engineer:

One popular role or career in data science, machine learning engineer, is primarily responsible for automating data analysis processes. They design and implement machine learning systems, research, and optimize machine learning algorithms, and perform machine learning tests to monitor system performance and functioning. It has emerged as one of the most essential data science careers in recent times. (Sharma)

Database Administrator:

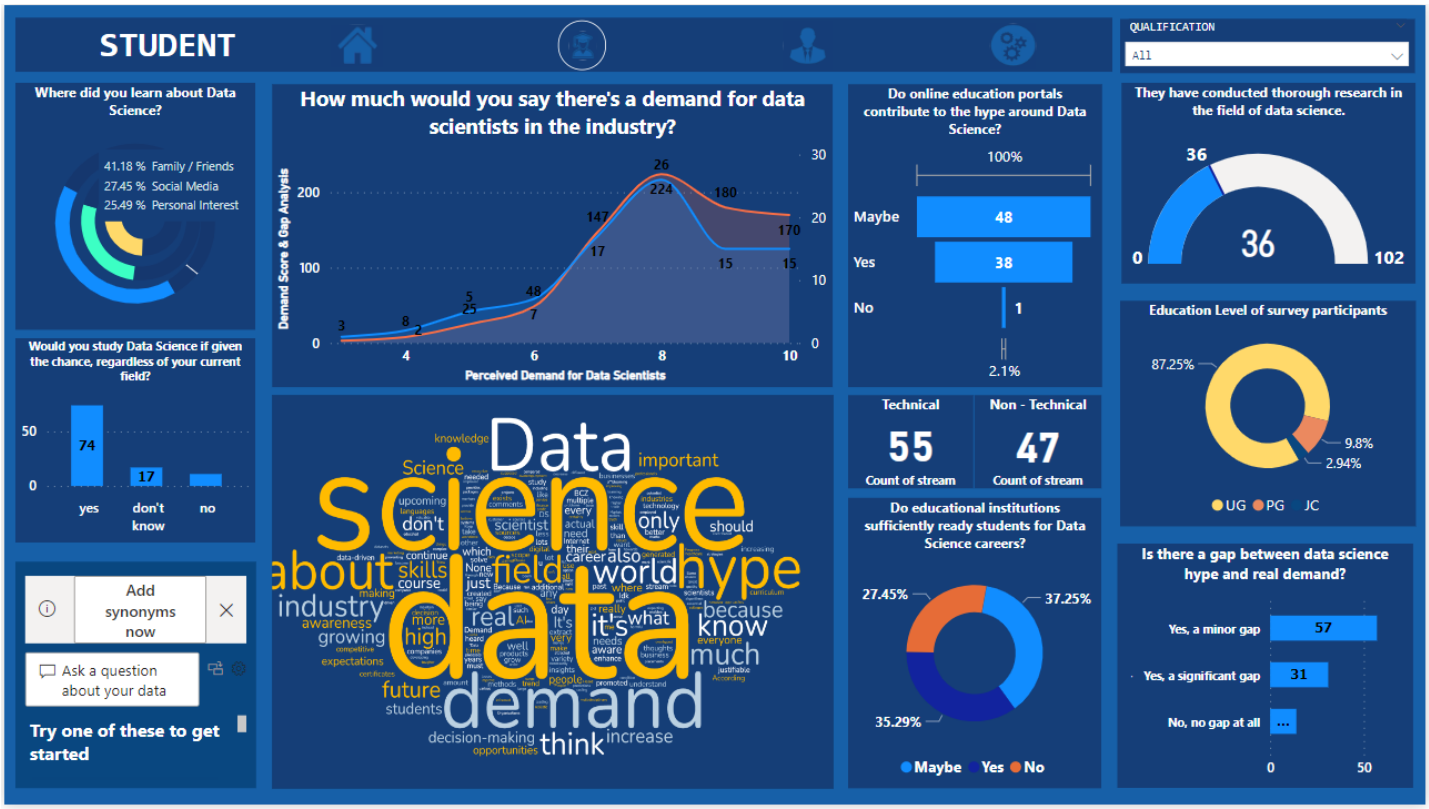
A database administrator is involved in maintaining and functioning the organization's database system. They are responsible for the safe and secure management and storage of data and develop backups and recovery solutions. To put it simply, they are responsible for the day-to-day functioning and management of a company's database system. (Sharma)

V. Findings -

[Student]

1. **Educational Background of participants:** **97.05%** of respondents had at least undergraduate or postgraduate degrees, indicating the data primarily reflects the perspectives of individuals with higher education.

2. **Personal Connections Driven Data Science Exploration:** The study found that individuals learn about Data Science primarily through **family/friends (41.18%)**, **social media (27.45%)**, and **personal interest (25.49%)**. This indicates the vital role of personal connections and individual curiosity in fostering initial exposure to the field.



3. **Strong Demand for Data Science Skills:**
   A significant **72.54%** of respondents expressed willingness to study Data Science regardless of their current educational background given a chance. This shows the perceived value of this field.

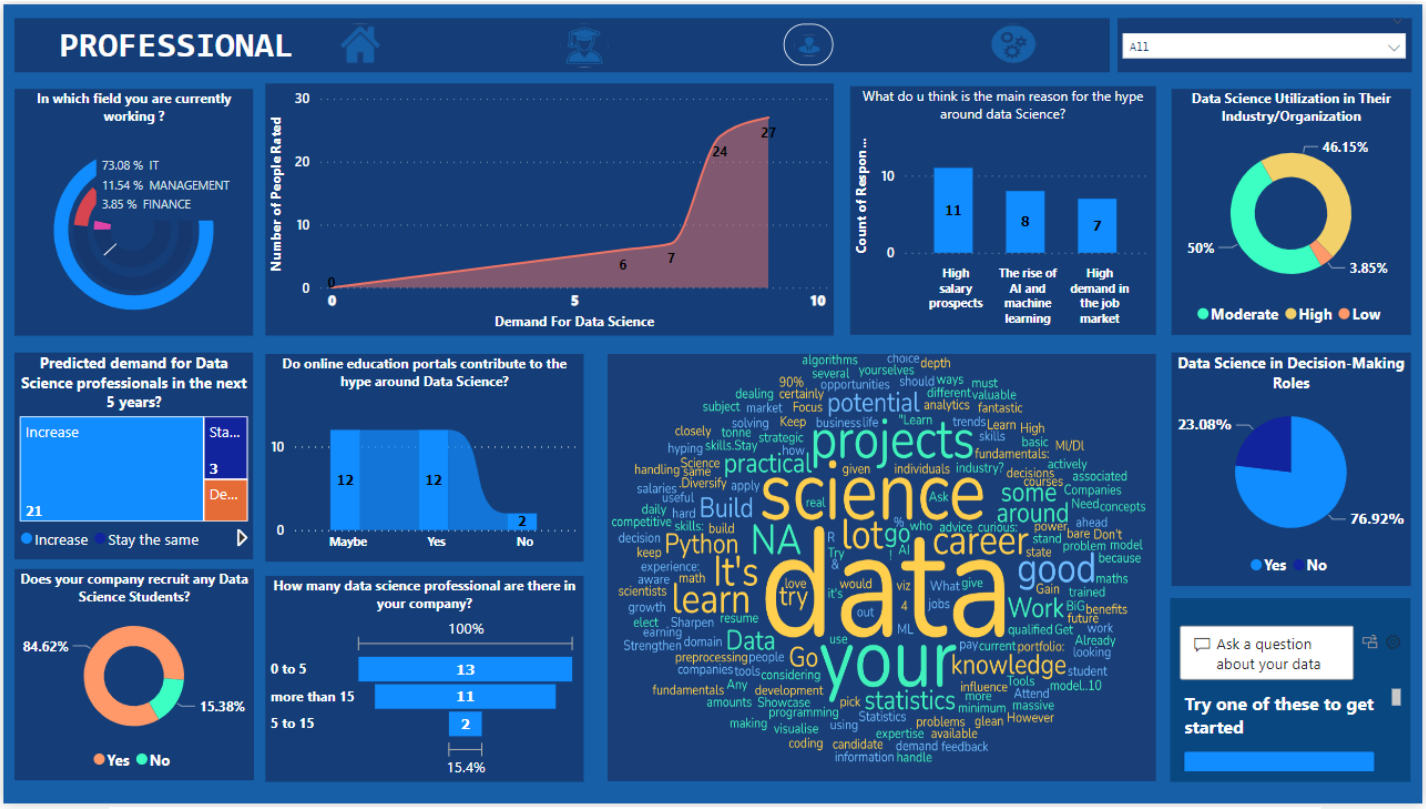
4. **Online educational portals and Perceived hype:**
   Online education portals may contribute to the hype around Data Science, with **38%** of respondents indicating that they do. However, a significant portion of respondents **(48%)** are unsure or have no opinion on the matter.

5. **Addressing the disparity between the Hype and Demand gap:**
   A notable **64.70%** of survey participants acknowledged a disparity between the hype surrounding Data Science and the actual reality, and within this group, **37.25%** expressed that the gap was particularly significant. These findings imply a possible overestimation of job opportunities or unrealistic expectations regarding careers in Data Science.

6. **Closing the Skills Gap:**
   While **27.45%** of respondents felt a minor gap between educational programs and job readiness, this highlights the potential need for further development of Data Science curricula within educational institutions to better prepare graduates for the workforce.

7. (Professionals)**Importance of Both Technical and Non-Technical Skills:**

While **55%** identified technical skills as crucial, **47%** emphasized the importance of non-technical skills like communication, business acumen, and decision-making. This finding underscores the need for a comprehensive skillset encompassing both technical and non-technical expertise for successful data science careers.



[Professional]

Among the professional participants, **73.08%** were IT professionals, **11.54%** were Management professionals and the remaining **3.85%** were Finance professionals.

1. **Anticipated growth in the field of Data Science:**
The study reveals significant optimism surrounding the outlook of Data Science career prospects, with **80.76%** of respondents anticipating a rise in demand within the next five years**.** This aligns with the increasing need for data-driven insights across various industries.

2. **Companies' commitment to hiring Data Science students:**
84.62% of companies plan to increase or maintain recruitment efforts for Data Science students, signifying a strong commitment to building their data science capabilities. This finding underscores the critical role data science plays in today's competitive landscape.

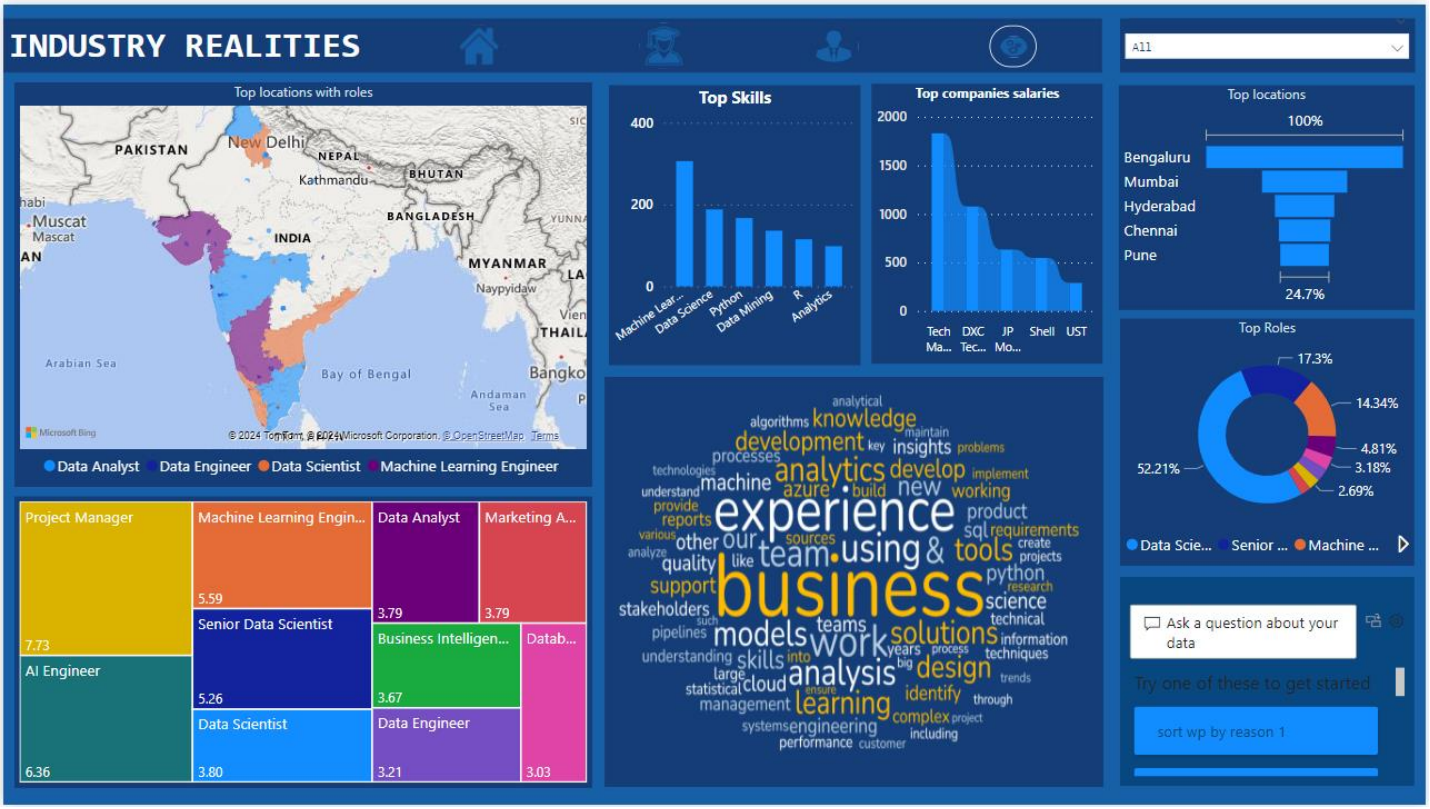
3. **Online Learning's Potential Impact on Hype:**

The study investigates the perceived influence of online education portals **46.15%** of respondents believe these platforms contribute to the hype surrounding Data Science, raising potential concerns about unrealistic expectations.

4. **Reasons for Data Science Hype:** Respondents mentioned several reasons for the hype around Data Science, including high salary prospects, the rise of AI and machine learning, and the need for advanced analytics skills for mitigating modern-day business problems.

5. **Moderate to High Data Science Utilization:** **50%** of respondents reported moderate Data Science usage, while **46.15%** indicated high usage within their respective industries/ organizations. This indicates a growing, but varied, integration of Data Science across different sectors.

6. **Utilization of Data Science in Decision-Making:** **76.92%** of respondents utilize data science in decision-making roles, while **23.08%** do not. This highlights the significant integration of data science in supporting informed decision-making processes across various organizations and industries.



[Industry Realities]

I. **Data Science Opportunities in Indian Cities:**

Mumbai, Hyderabad, Bengaluru, Pune, and Gurgaon are the top five Indian cities offering data science job opportunities. Bengaluru has the highest percentage and Lucknow has the lowest percentage of data science job opportunities.

II. Experience scale in Data Science Roles:

The highest average experience requirement for a "Project Manager" is **7.73 years**, indicating it's a senior position. "AI Engineer" and "Senior Data Scientist" follow with **6.36** and **5.26 years**, respectively.

III. Experience Tiers for Data Science Positions:

Entry to mid-level positions includes "Data Analyst", "Business Intelligence Analyst," and "Data Engineer" at **3.79**, **3.67**, and **3.21** minimum average experience. "Marketing Analyst" and "Database Administrator" are entry-level, needing **3.79** and **3.03** minimum average years.

IV. In-Demand Skills for Data Science Roles:

The data suggests that **Machine learning, Python, Data mining, R analytics, and NLP (Natural Language Processing)** are some of the most in-demand skills in the field of Data science and analytics. These skills are essential for Data Science Professionals to have to be successful in their roles.

V. Insights from Data Professionals:

The word cloud represents the views of data science professionals on data science. It includes terms such as machine learning, data analysis, algorithms, Python, Azure, and project management, indicating the importance of these concepts in data science. Other terms like stakeholders, business, and performance suggest the focus on delivering value and meeting business objectives in data science projects.

The word cloud emphasizes the interdisciplinary nature of data science, combining aspects of statistics, machine learning, software development, and project management.

VI. Problem Statements: -

- Significant gap between the hype of data science and actual job realities
- Overestimation of job opportunities and prospects
- Deficiencies in data science curricula, not fully preparing for jobs.
- Continuous change in the data science job market
- Uncertainty about the longevity of data science roles and skills in the future job market
- Jobs are more concentrated in the Metropolitan cities
- It takes a significant amount of time to become a Data Science professional as compared to other IT professions.

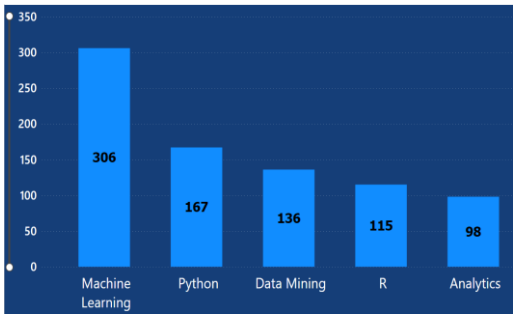
VII. Solutions: -

- To close the gap between the expectations and actual job realities partner with industry leaders to shape data science education that reflects real-world needs. Let practicing data scientists share their experiences to bridge the gap between perception and reality.

- Develop platforms offering transparent data on job markets, including demand trends and success stories, to provide a clearer picture of data science opportunities. Base career guidance on thorough market research to give students accurate information about job availability and growth sectors. Highlight the diverse applications of data science across various domains to showcase its broad impact and potential career paths.

- Programs should include an adaptable skillset to cultivate a continuous learning mindset. Equipping professionals to stay relevant amidst evolving technologies.

Offer back-to-work and bridging programs to help professionals re-enter the field after a hiatus. As per our data following are the essential skills for a Data Science professional as long as there is no breakthrough in the IT industry.

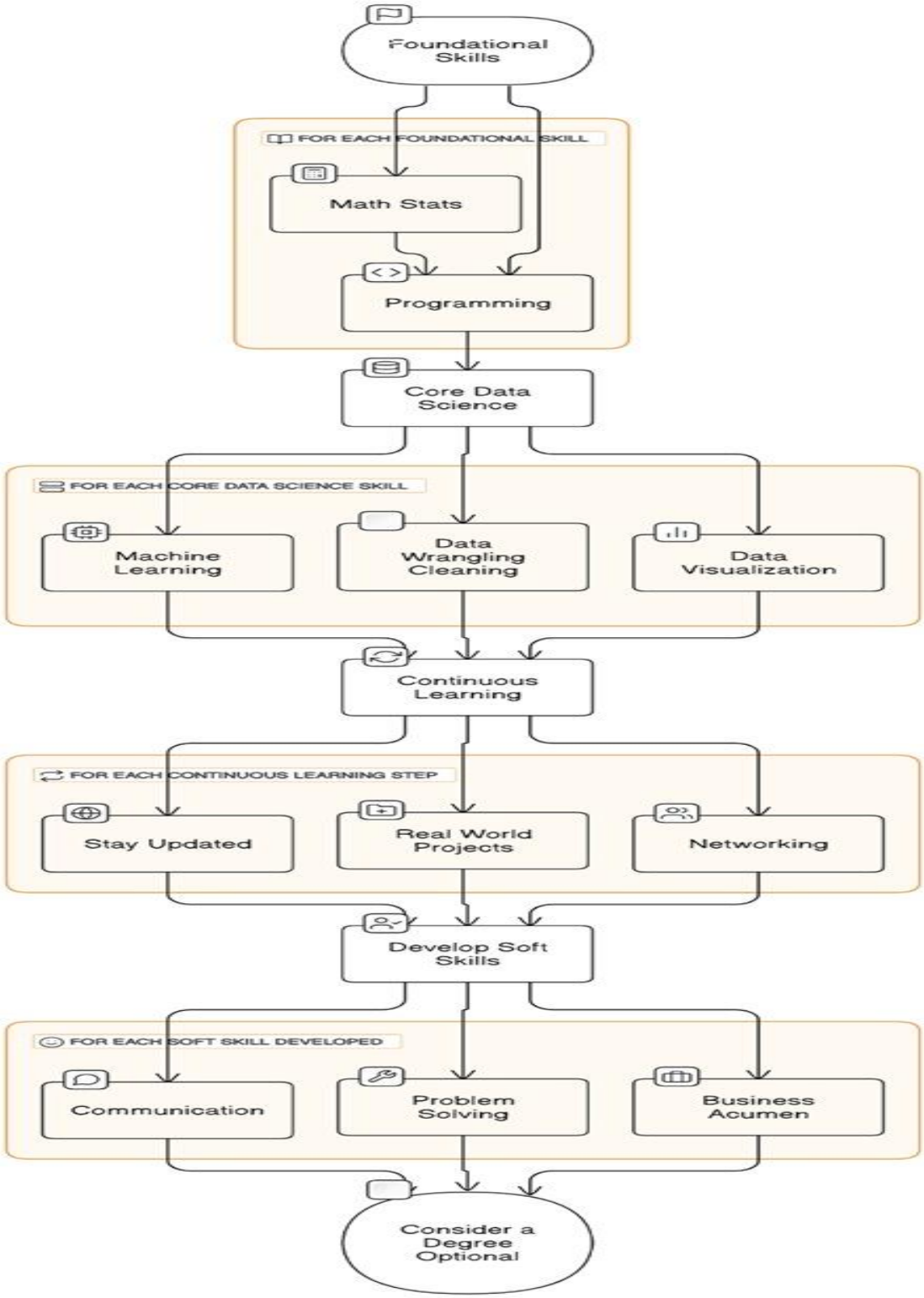


- Encourage and facilitate remote work policies within the data science industry. This allows companies to tap into a broader talent pool, including individuals outside metropolitan areas. Implement skill development programs in collaboration with local governments and educational institutions to equip individuals in non-metropolitan areas with the necessary skills for data science roles.

- To navigate the substantial time investment required to attain proficiency in Data Science, individuals should possess AI tools and adept prompt engineering skills. These resources are instrumental in diminishing the time required to achieve excellence in the field of data science.

- To overcome the deficiencies and lacking in the data science curriculum taught in the institutes infuse data science programs with industry insights to ensure they remain relevant. Integrate mandatory real-world projects and internships for the practical application of knowledge. Broaden the focus beyond technical skills to encompass communication and business acumen. Following is a roadmap to becoming a Data Science Professional.

Roadmap to Become a Data Science Professional



VIII. Conclusion:-

The field of data science has undoubtedly garnered immense hype and attention, fueled by its perceived potential as a lucrative and impactful career choice. However, as this study has revealed, a significant gap exists between the heightened expectations surrounding data science and the pragmatic realities of its demand within the job market.

Through comprehensive data analysis and insights from both students and professionals, this research has unveiled critical factors contributing to this disparity. Unrealistic expectations, coupled with an overestimation of job opportunities, including the rise of AI and machine learning, the need for advanced analytics skills, and the allure of high salary prospects. However, the study reveals that online educational portals may also play a role in perpetuating the hype, potentially leading to misconceptions about job availability and career paths.

Nonetheless, the findings underscore the sustained demand for data science skills across various industries, with a majority anticipating growth in this domain within the next five years. However, the study also highlights the need for continuous adaptation and skill development to navigate the ever-evolving data science job market.

Furthermore, the concentration of Data Science opportunities in metropolitan cities and the substantial time investment required to attain proficiency pose challenges that need to be addressed through remote work policies, skill development programs, and the integration of AI tools and prompt engineering.

By fostering collaboration between industry leaders and educational institutions, implementing transparent job market data platforms, and creating an effective roadmap to becoming a Data Science professional that reflects real-world needs, the study proposes solutions to align expectations with realities.

Furthermore, emphasizing a comprehensive skill set encompassing both technical and non-technical competencies is crucial for success in data science roles. Adaptability, continuous learning, and the integration of practical experiences through internships and projects should be prioritized within educational programs.

As the data science landscape continues to evolve, this study serves as a catalyst for informed decision-making and strategic career planning. By bridging the gap between hype and industry realities, aspiring students and professionals can navigate the complexities of the data science job market with greater clarity and preparedness.

IX. Github Dashboard Link:

Aavishkar-Project/Aavishkar Dashboard(In Progress).pbix at main · Tempest-Kohaku/Aavishkar-Project (github.com)

X. Reference: -

References

Bhadoria, Vikas. *data-scientist-role-in2020*. . . 2020. csv. 1 January 2024.
<https://www.kaggle.com/datasets/vikasbhadoria/data-scientist-role-in2020>.

Duarte, Fabio. *Amount of Data Created Daily (2023)*. 3 April 2023. <https://explodingtopics.com/blog/data-generated-per-day>.

Hale, Jeff. *data-scientist-general-skills-2018-revised*. . . 2019. csv. 1 January 2024. <https://www.kaggle.com/datasets/discdiver/data-scientist-general-skills-2018-revised>.

Pant, Madhur. *data-science-jobs-in-india*. . . 2022. csv. 1 January 2024. <https://www.kaggle.com/datasets/madhurpant/data-science-jobs-in-india>.

Professionals, Students &. *Survey on DS Opinion* TCSC DS Avishkar Team. 1 August 2023. Google Sheet, Google Meet. <https://docs.google.com/spreadsheets/d/1CqFdQqh19vwy1dwN9dCrk6aHaHR9XWyPq-AE9mYdFLw/edit?usp=sharing>.

Sharma, Rohit. *upGrad*. 29 August 2022. MLA. 7 December 2023. <https://www.upgrad.com/blog/career-in-data-science/>.