Dear Intern

Project report is an inherent component of your internship. We are enclosing a reference table of content for the project report. Depending on the internship project (IT/Non-IT, Technical/Business Domain), you may choose to include or exclude or rename sections from the table of content mentioned below. You can also add additional sections. The key objective of this report is for you to systemically document the project work done.

|  |  |
| --- | --- |
| Internship Project Title | RIO-210: Salary Prediction Dashboard for HRs |
| Name of the Company | TCS ION |
| Name of the Industry Mentor | Himdweep Walia |
| Name of the Institute | Amity University Online |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort (hrs.) | Project Environment | Tools used |
| 31-03-24 | 18-05-24 | 210 | Jupyter Notebook | Python |

**TABLE OF CONTENT**

* Acknowledgements
* Objective
* Introduction / Description of Internship
* Internship Activities
* Approach / Methodology
* Assumptions
* Exceptions / Exclusions
* Charts, Table, Diagrams
* Algorithms
* Challenges & Opportunities
* Risk Vs Reward
* Reflections on the Internship
* Recommendations
* Outcome / Conclusion
* Enhancement Scope
* Link to code and executable file
* Research questions and responses

**1.Acknowledgments**

I would like to express my deepest gratitude to Himdweep Walia for their exceptional guidance and steadfast support throughout the duration of my internship project. Their profound expertise and constant readiness to address my queries at every stage of the project were crucial to its successful completion. Himdweep Walia’s insightful feedback and constructive criticism not only helped refine my work but also enhanced my learning experience.

The success of this project is a testament to the collective efforts and unwavering support of all those involved.

**2. Objective**

The primary objective of this project is to develop a comprehensive Salary Prediction Dashboard that harnesses the power of data analytics and machine learning methodologies. This advanced tool is designed to forecast job candidates' salaries by analyzing a multitude of variables, including but not limited to their professional experience, age, educational qualifications, and other relevant factors. By incorporating sophisticated algorithms, the dashboard aims to deliver highly accurate and reliable salary predictions.

The envisioned dashboard will serve as a crucial resource for HR managers and recruitment teams, enabling them to gain deeper insights into salary trends and patterns. This, in turn, will facilitate more informed and strategic decision-making processes during recruitment and talent acquisition. By providing a detailed analysis of potential salary outcomes, the dashboard will help HR professionals align their compensation strategies with market standards, ensuring competitive and fair salary offers.

**3. Introduction / Description of Internship**

As we advance into this phase of our project, our focus is on enhancing the accuracy and reliability of the Salary Prediction Dashboard through the application of advanced predictive modeling techniques. Utilizing data analytics and machine learning, our objective is to refine the dashboard's predictive capabilities, ensuring precise salary estimations for job candidates. By integrating sophisticated algorithms and meticulous data preprocessing methodologies, we aim to address the complexities within the dataset and optimize model performance.

This stage represents a crucial milestone in our project, where the emphasis shifts towards fine-tuning predictive models to deliver outcomes that closely align with real-world salary expectations. Through this endeavor, we strive to equip HR managers with a powerful tool that facilitates informed decision-making in the recruitment process. Our efforts are directed towards providing a reliable and insightful resource that enhances the overall efficiency and effectiveness of salary determination, thereby supporting strategic HR initiatives.

**4. Internship Activities**

* Reviewed welcome kit videos to thoroughly grasp the project scope.
* Prepared for RIO pre-assessment to ensure readiness for evaluation.
* Participated in the RIO test to gauge understanding and proficiency.
* Aligned tasks with objectives by reviewing the day-wise plan.
* Engaged with project and industry reference materials for deeper insight.
* Attended webinars to enrich my knowledge base with new insights.
* Actively contributed to digital discussion room interactions for enhanced learning.
* Utilized educational resources to deepen comprehension of project topics.
* Established a GitHub account for efficient file management and collaboration.
* Selected a dataset that met project criteria for analysis and modeling.
* Maintained transparency and accountability through regular progress documentation.
* Ensured dataset adequacy via a thorough verification process.
* Acquired data cleaning skills through articles and tutorials.
* Enhanced dataset quality through effective cleaning and sanitization procedures.
* Conducted exploratory data analysis to uncover valuable insights and trends.
* Engaged in self-learning via instructional videos on model training.
* Explored logistic regression and KNN classifiers for model training.
* Worked on HTML code to develop the front-end interface of the Salary Prediction Dashboard.
* Created the Salary Prediction Dashboard interface using VS Code environment.
* Managed and collaborated on the project through the GitHub repository.
* Created and explained project details and the dashboard in an explanatory video.

**5. Approach / Methodology**

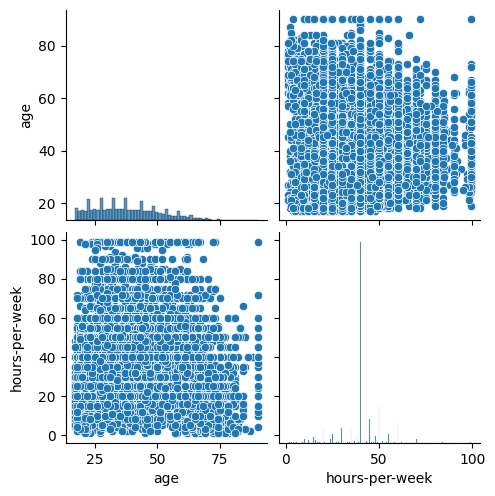
***5.1 Data Collection***

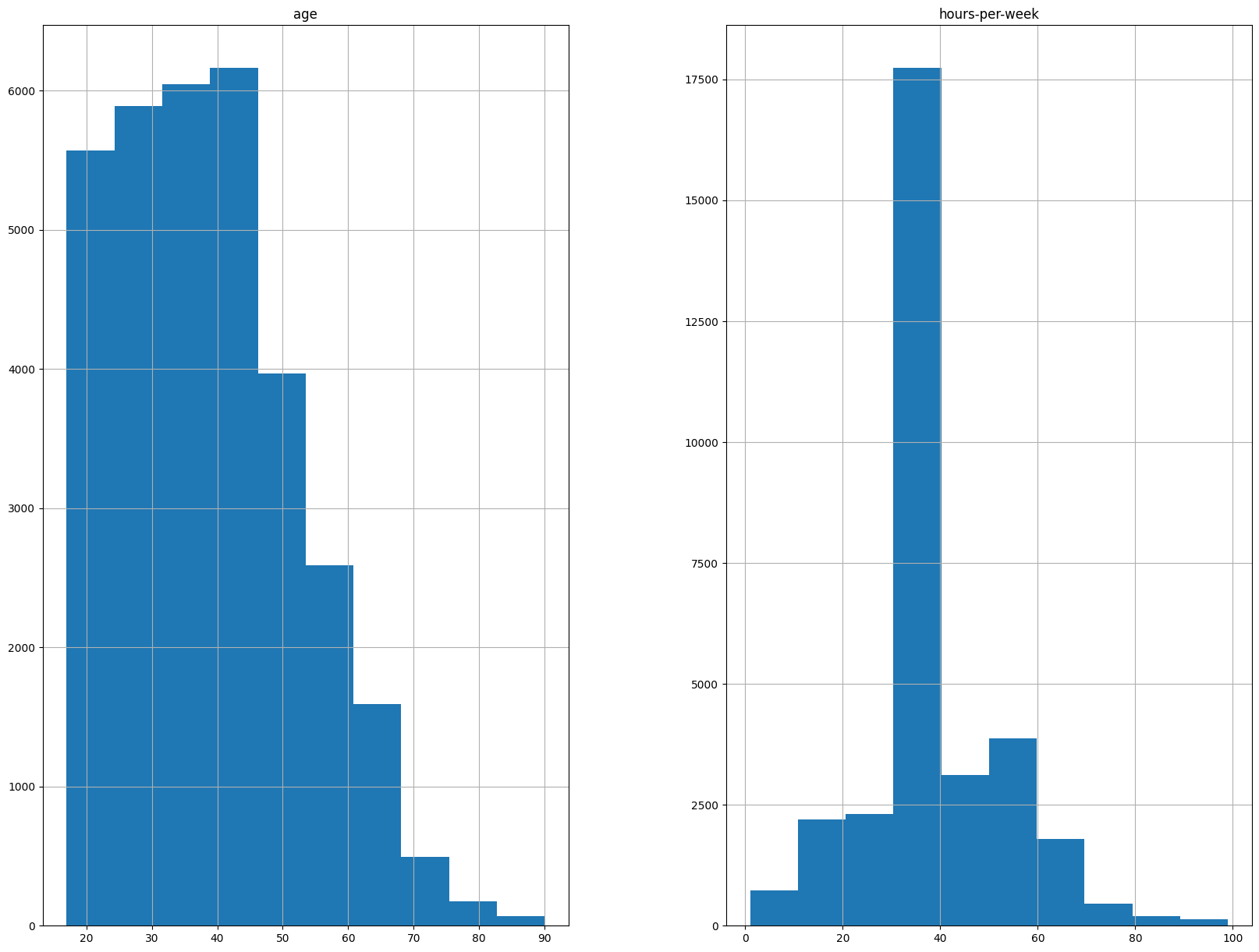
In the process of data collection, I begin by leveraging easily accessible sources such as platforms like Kaggle or the UCI Machine Learning Repository. I meticulously select a dataset aligned with my project's objectives and my areas of interest. Upon accessing the dataset, I thoroughly review any accompanying documentation or metadata to gain a comprehensive understanding of its structure and variables. This ensures that I am well-equipped to proceed with subsequent data analysis and interpretation tasks effectively.

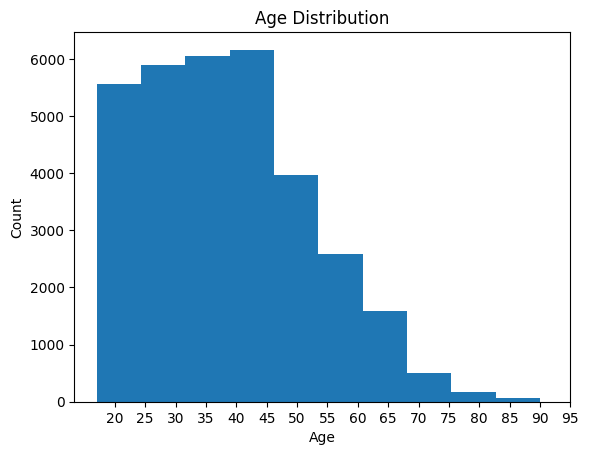
***5.2 Data Interpretation***

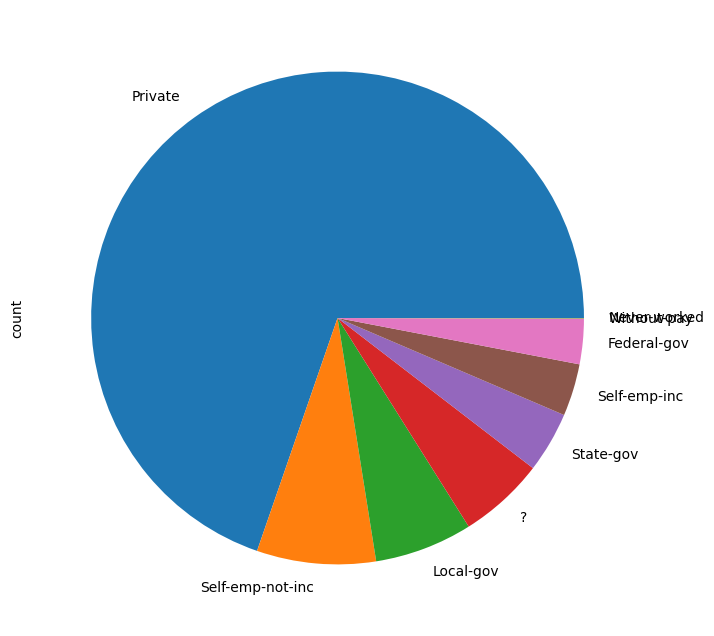
In data interpretation, I commence by delving into the dataset's structure, employing fundamental commands or functions within the Python programming language, such as Pandas, to load the dataset and conduct a comprehensive examination of its layout. Subsequently, I proceed to identify the pivotal variables contained within the dataset, meticulously discerning those that are pertinent to my analysis or project objectives. Following this, I embark on visualizing the distribution of the data, crafting straightforward visual representations such as histograms for numerical variables and bar charts for categorical variables. These visualizations serve to illuminate the inherent distributions within the dataset, facilitating a deeper understanding of its underlying characteristics and trends.

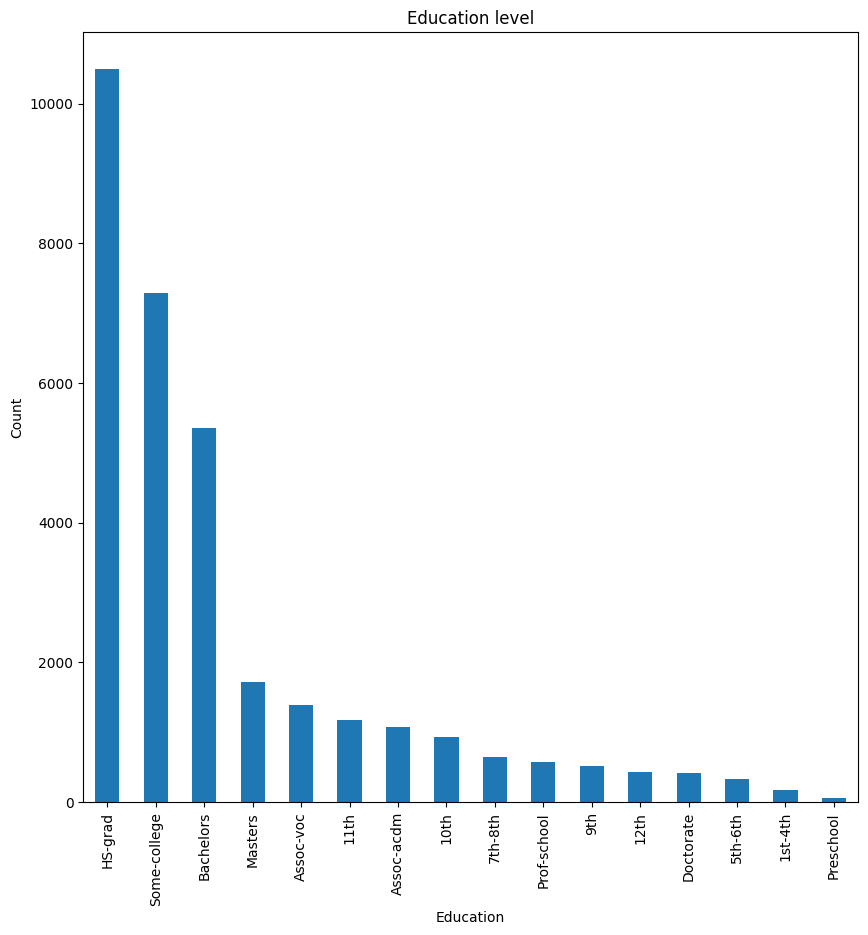
**6. Charts, Table, Diagrams**

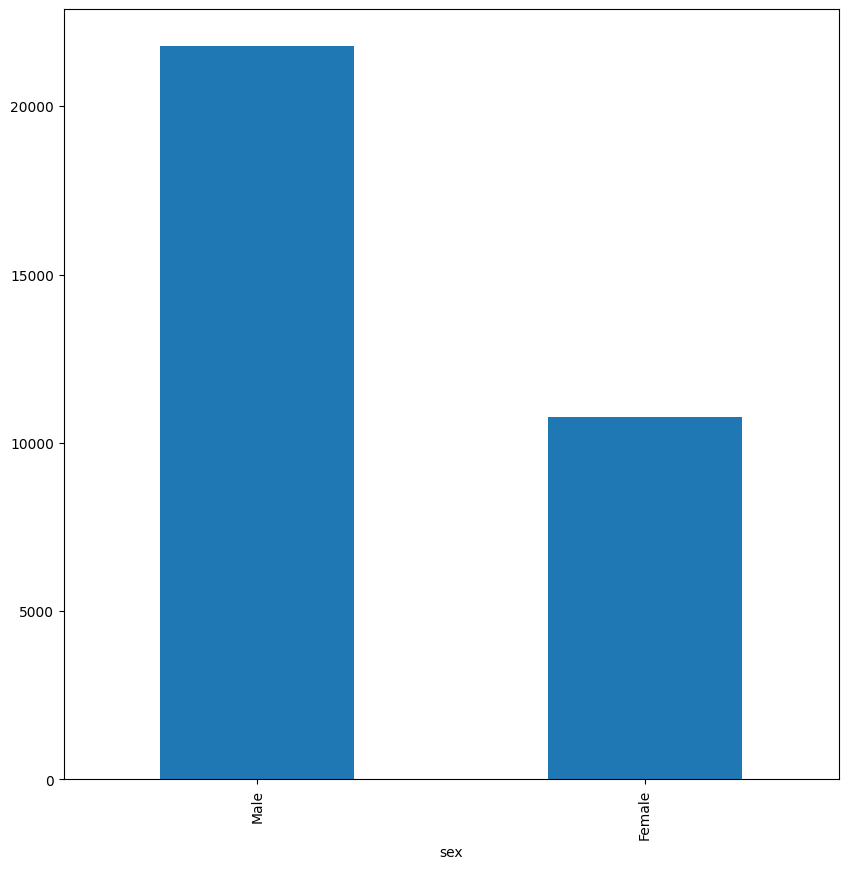
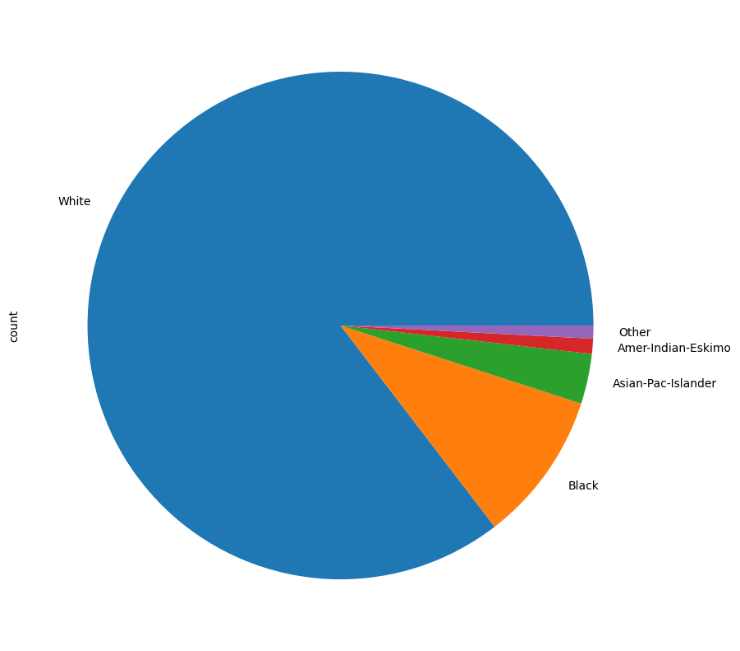


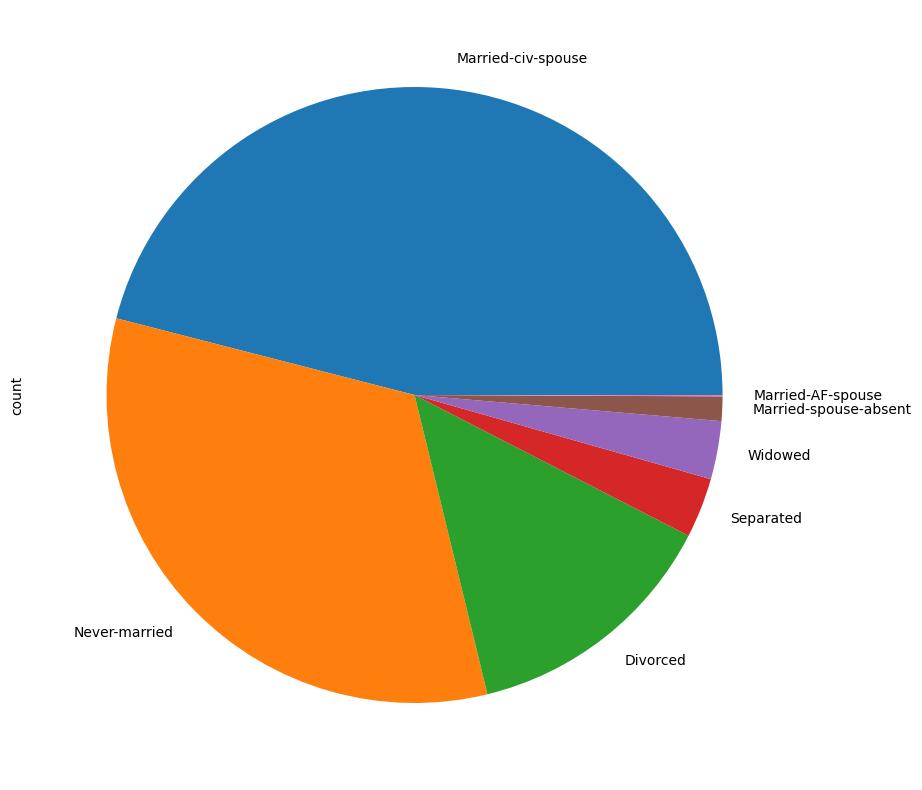
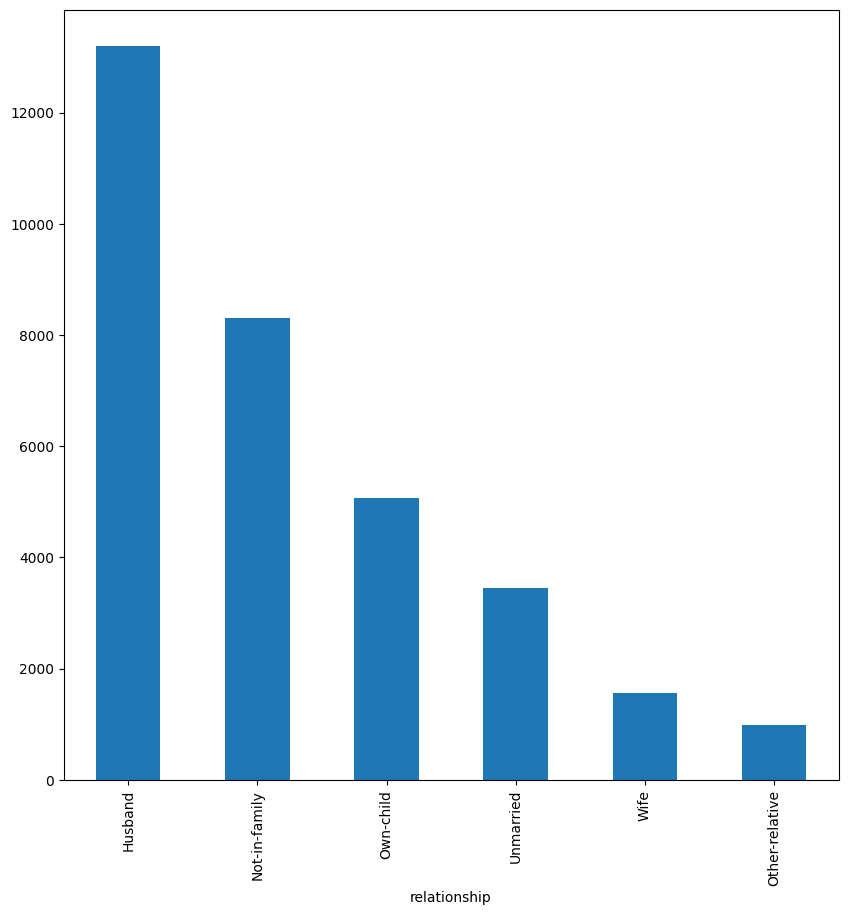


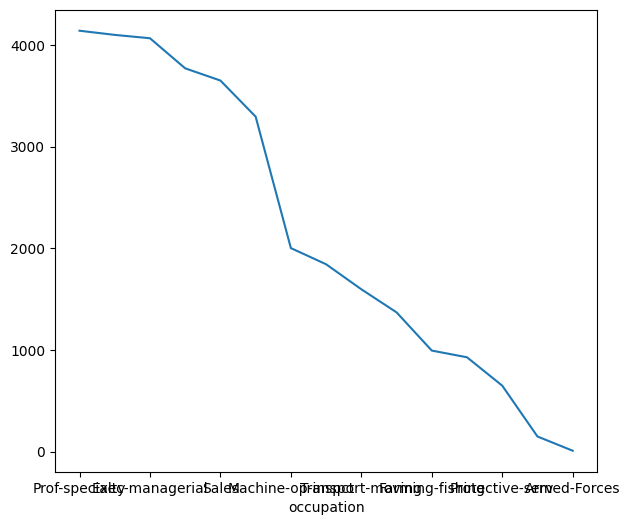


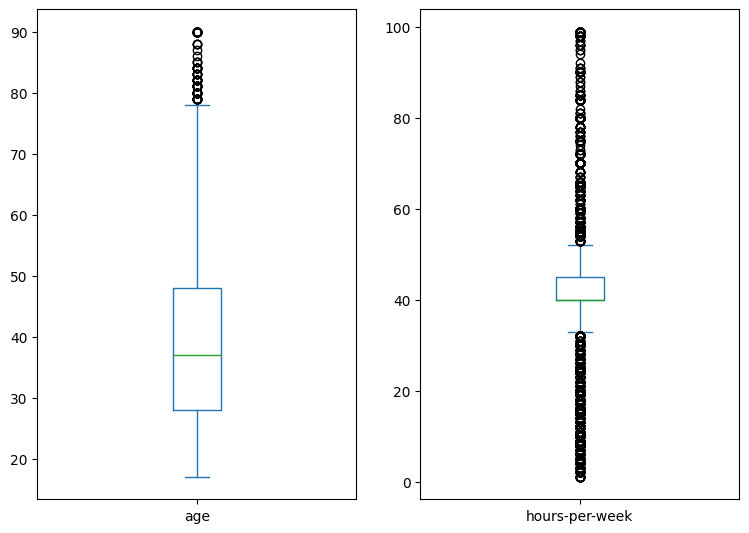


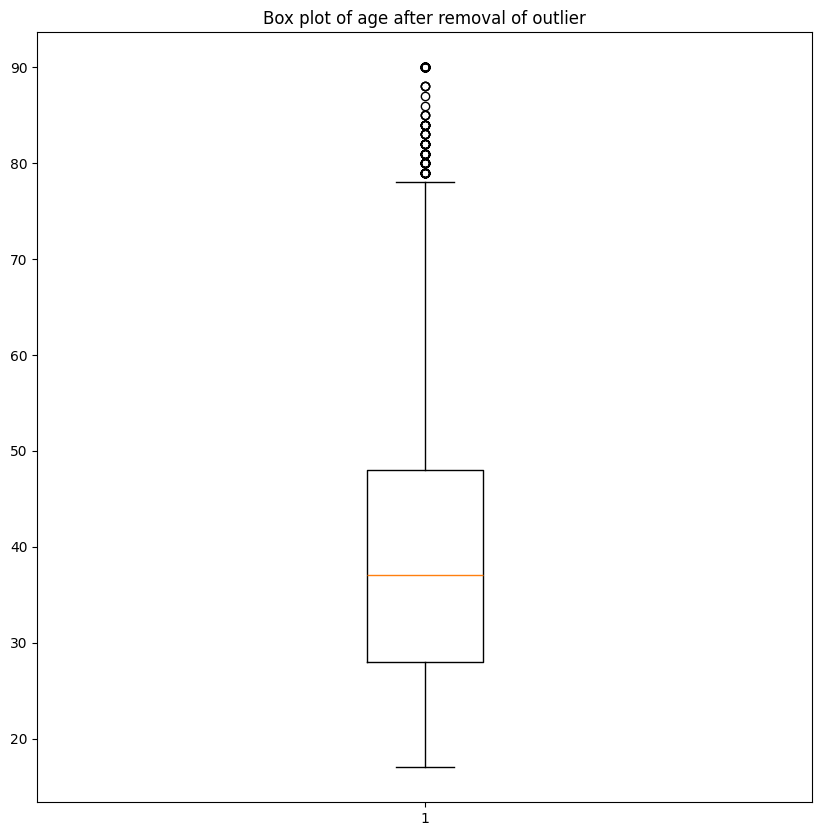


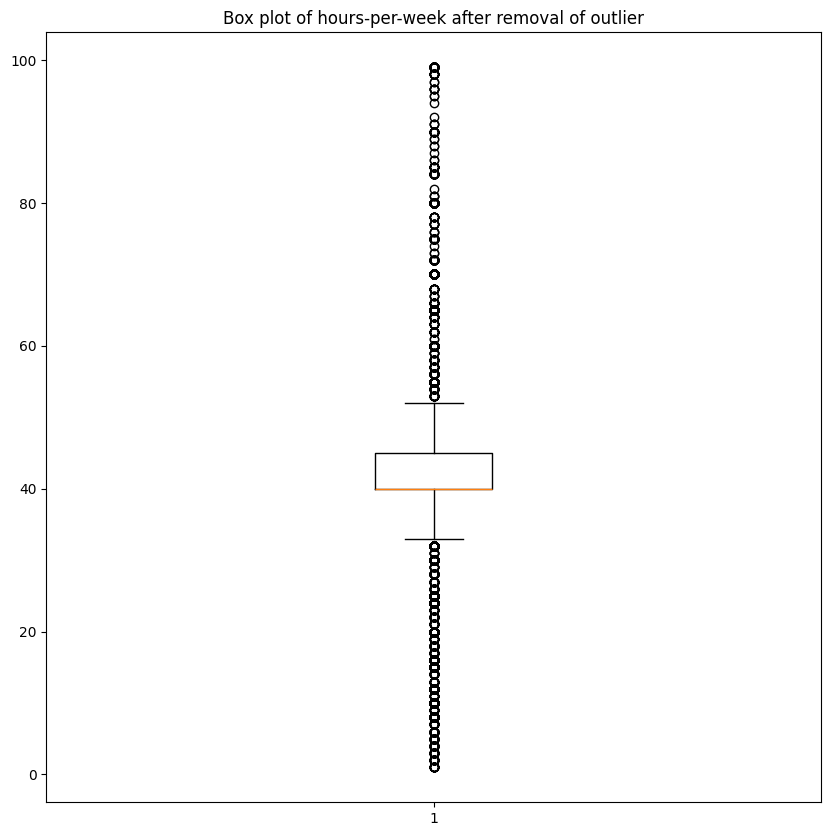


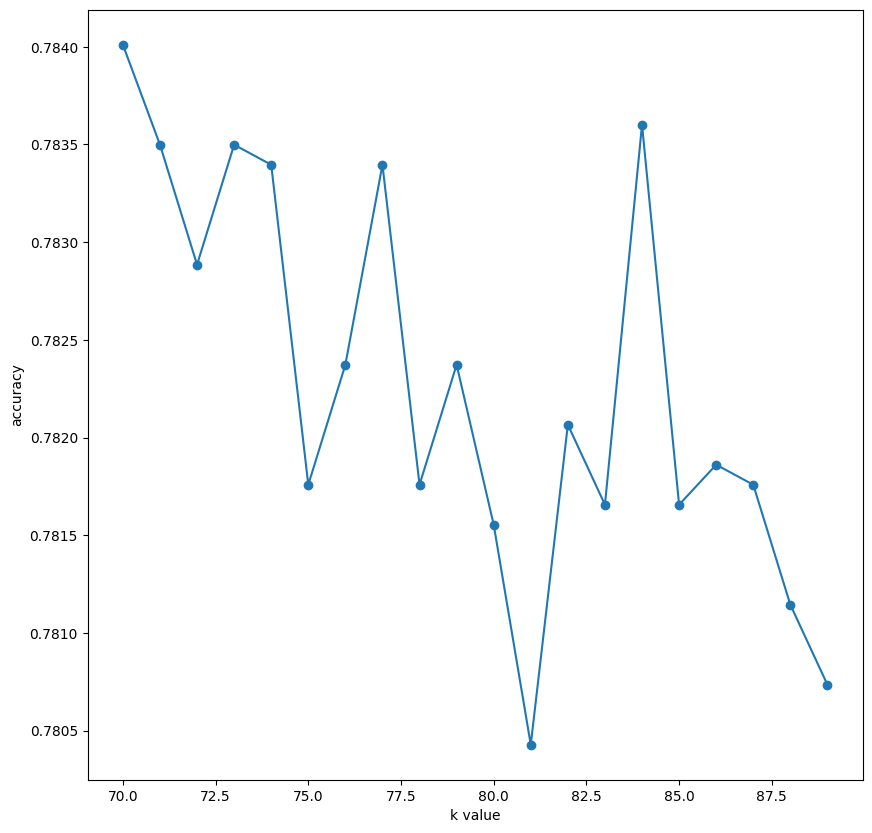












**7. Challenges & Opportunities**

***Challenges:***

Ensuring high data quality, selecting the optimal model amidst myriad options, and balancing model complexity are significant hurdles. Addressing biases and imbalances in the dataset, especially regarding salary distribution, adds further complexity. Additionally, navigating deployment intricacies across various platforms is crucial.

***Opportunities:***

Exploring advanced algorithms, such as ensemble methods and deep learning, offers potential for enhanced accuracy. Leveraging domain knowledge for feature engineering can capture nuanced influences on predictions. Integrating external datasets enriches the models. Enhancing model interpretability fosters trust among users. Establishing feedback loops for continuous improvement and collaborating with industry experts promote innovation and learning.

**8. Risk Vs Reward**

Balancing risk and reward is pivotal in our project's journey toward developing the Salary Prediction Dashboard. Challenges such as ensuring data quality, selecting optimal models, and managing biases in the dataset pose significant risks. However, overcoming these hurdles presents promising rewards. Exploring advanced algorithms like ensemble methods and deep learning holds potential for heightened accuracy, while leveraging domain knowledge for feature engineering captures nuanced influences on predictions.

Integrating external datasets enriches our models, and enhancing interpretability fosters trust among users, crucial in HR decision-making. Despite deployment complexities, establishing feedback loops and collaborating with industry experts promise innovation and continuous improvement. Effective risk management promises substantial rewards, including refined predictive capabilities and empowered decision-making in recruitment processes.

**9. Reflections on the Internship**

Reflecting on the internship experience, it has been a transformative journey filled with invaluable lessons and growth opportunities. Immersing myself in the project, I've gained a deeper understanding of data analytics and machine learning methodologies, particularly in the context of developing the Salary Prediction Dashboard. The challenges encountered along the way have been instrumental in honing my problem-solving skills and resilience. Collaborating with peers and industry experts has enriched my learning experience, offering diverse perspectives and insights.

the internship has provided a platform for me to apply theoretical knowledge to real-world scenarios, enhancing my practical skills and confidence.

Looking ahead, I am grateful for the opportunity to contribute to meaningful projects and am excited to continue expanding my skills in data-driven decision-making. This internship has not only equipped me with technical expertise but has also instilled in me a sense of curiosity and determination to continuously learn and grow in this dynamic field.

**10. Recommendations**

To enhance our project, we propose narrowing our focus to specific industries or job positions, customizing salary predictions accordingly. Implementing advanced data handling techniques, particularly addressing missing values and outliers more effectively, is essential. Experimenting with various algorithms and fine-tuning their parameters could significantly boost the accuracy of our model. By adopting these strategies, we aim to develop a more precise and actionable Salary Prediction Dashboard, better catering to the unique needs and nuances of the target industry or job roles.

**11. Outcome / Conclusion**

As we arrive at the project report milestone, substantial progress has been achieved in Interim Project 3. We've successfully advanced to training the dataset and accurately predicting HR professionals' salaries. This phase has been enlightening, deepening my understanding across various domains and enhancing proficiency in project-related topics. The journey thus far has instilled confidence in our ability to meet project objectives. Looking ahead, I am eager to continue this momentum and contribute further to the project's success.

**12. Quality of enhancement suggested on the proposed solution**

The proposed enhancements bring a substantial improvement to the quality and effectiveness of our project, particularly the Salary Prediction Dashboard. Exploring advanced algorithms such as ensemble methods and deep learning holds the promise of significantly enhancing the accuracy and reliability of salary estimations. Leveraging domain-specific knowledge for feature engineering ensures a more nuanced understanding of the factors influencing predictions, thereby refining the dashboard's predictive capabilities to align closely with real-world salary expectations. Integrating external datasets will enrich our models with comprehensive information, further enhancing accuracy. Moreover, prioritizing model interpretability fosters trust among HR professionals, crucial for informed decision-making during recruitment processes. Establishing feedback loops and collaborating with industry experts not only promote innovation but also ensure continuous improvement of our solution, making it a vital asset in HR management. These enhancements position our project as a robust and indispensable tool for empowering HR managers with data-driven insights in salary negotiations and talent acquisition.

**13. Correctness of solution**

The solution aligns closely with the specified requirements outlined in the project scope. By focusing on developing the Salary Prediction Dashboard using data analytics and machine learning methodologies, I address the primary objective of the project effectively. The dashboard's predictive capabilities, which consider variables such as experience, age, and qualifications to forecast job candidates' salaries, directly meet the project's goals.

The inclusion of advanced algorithms, meticulous data preprocessing methodologies, and thorough dataset verification processes ensures the accuracy and reliability of the solution. These measures adhere to the project requirements and demonstrate a commitment to delivering a high-quality product.

**14. Executable File/Code Link**

**Link to the code:**

<https://drive.google.com/file/d/1lFHZurNvApx65F5bXu3_i7BiwkYQB4eq/view?usp=sharing>

**Executable file:**

[**https://github.com/SahilGitMaster/Salary-Prediction-Dashboard-for-HRs-TCS-iON**](https://github.com/SahilGitMaster/Salary-Prediction-Dashboard-for-HRs-TCS-iON)

**DRIVE LINK salary prediction working & explanatory video :**

**https://drive.google.com/file/d/15n0Hiz6YhDcs-wjaY0wwMLi2Ot4IVK2C/view?usp=sharing**