

បណ្ឌិត្យសភាបច្ចេកវិទ្យាឌីជីថលកម្ពុជា

មហាវិទ្យាល័យ វិទ្យាសាស្ត្រ កំព្យូទ័រ

គម្រោងសញ្ញាបត្របរិញ្ញាបត្រ

របស់និស្សិត៖ ម៉ាក់​ ​ច័ន្ទណា

កាលបរិច្ឆេទការពារនិក្ខេបបទ៖

អនុញ្ញាតអោយការពារគំរោង

ប្រធានបណ្ឌិត្យសភាបច្ចេកវិទ្យាឌីជីថលកម្ពុជា

ថ្ងៃទី ខែធ្នូ ឆ្នាំ២០២៤

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ប្រធានបទ៖​ Cambodia Job Trending and Insight from 2021 to 2023

សហគ្រាស៖ Institute of Digital Research and Innovation (IDRI)

គ្រូដឹកនាំគម្រោង៖ អ៊ូច​ គង្គារ​

អ្នកទទួលខុសត្រូវក្នុងសហគ្រាស៖ ធៀ​ សុផល​

រាជធានីភ្នំពេញ



**Cambodia Academy of Digital Technology**

**School of Computer Science**

**INTERSHIP REPORT OF**

Mr. Mak Channa

**Defense Date: Month Day year**

**Internship defense authorization**

**President: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Phnom Penh, December \_\_\_, 2023**

**Topic : Cambodia Job Trending and Insight from 2021 to 2023**

**Enterprise : Institute of Digital Research and Innovation (IDRI)**

**Advisor : Mr. Ouch Kongkea**

**Supervisor : Mr. Thear Sophal**

**PHNOM PENH**

**INTERNSHIP REPORT ON**

**Cambodia Job Trending and Insight from 2021 to 2023**

**at**



**Institute of Digital Research and Innovation (IDRI)**

*Submitted By:*

**Mak Channa**

*Under the supervision of:*

**Thear Sophal**

*Under the advisory of:*

**Ouch Kongkea**

Date of defense: Month Day Year

*In partial fulfillment of the requirements for the Internship Program*

*for the award of the degree of*

Bachelor of ****Computer Science



School of Digital Engineering

Cambodia Academy of Digital Technology

# ACKNOWLEDGEMENT

Initially, I would like to express my sincere gratitude to His Excellency President H.E. Dr. Seng Sopheap President of the Cambodia Academy of Digital Technology, Your vision and dedication have helped to make CADT a place where students can thrive and achieve their full potential.

Then, I would like to express my sincere gratitude to my advisor, Mr.Ouch Kongkea for his help and guidance on my project. His valuable insights and suggestions have helped me improve the quality of my work immensely. I am grateful for their willingness to take the time to help me, and I am confident that I would not have been able to complete this project without his assistance.

Eventually, I would particularly like to thank my supervisor, Mr. Thear Sophal for his mentorship and guidance throughout the internship. He also has provided me with the opportunity to learn and grow in my field, and I am grateful for his support. I am excited to use the skills and knowledge that I have gained to pursue a career in the future.

Finally, I would like to express my sincere gratitude to my family for their support and encouragement throughout the completion of this report project. My parents have always been my biggest supporters, and I am grateful for their guidance and advice. I am also grateful to my friends for their support and understanding during this time. I could not have completed this project without the help and support of my family and friends

# ABSTRACT

# មូលន័យសង្ខេប

# TABLE OF CONTETNS

[ACKNOWLEDGEMENT i](#_Toc161317424)

[ABSTRACT ii](#_Toc161317425)

[មូលន័យសង្ខេប i](#_Toc161317426)

[TABLE OF CONTETNS ii](#_Toc161317427)

[LIST OF FIGURES i](#_Toc161317428)

[LIST OF TABLES ii](#_Toc161317429)

[LIST OF ABBREVIATIONS iii](#_Toc161317430)

[I. INTRODUCTION 1](#_Toc161317431)

[1.1. Presentation of Internship 1](#_Toc161317432)

[1.1.1. Objective of Internship 1](#_Toc161317433)

[1.1.2. Duration of Internship 1](#_Toc161317434)

[1.2. Presentation of Organization 1](#_Toc161317435)

[1.2.1. General Information of Company 1](#_Toc161317436)

[1.2.2. Services of Company 2](#_Toc161317437)

[1.2.3. Vison & Mission 2](#_Toc161317438)

[1.2.4. Address & Contact 3](#_Toc161317439)

[II. PRESENTATION OF THE PROJECT 4](#_Toc161317440)

[2.1. General Presentation of Project 4](#_Toc161317441)

[2.2. Problematic 4](#_Toc161317442)

[2.3. Objective 4](#_Toc161317443)

[III. LIERATURE REVIEW 5](#_Toc161317444)

[IV. PROJECT ANALYSIS AND CONCEPTS 5](#_Toc161317445)

[4.1. Tools 5](#_Toc161317446)

[4.2. Technology requirement 6](#_Toc161317447)

[V. METHODOLOGY 7](#_Toc161317448)

[5.1. Project Pipeline 7](#_Toc161317449)

[5.2. Overview data 7](#_Toc161317450)

[5.3. Data preprocessing 7](#_Toc161317451)

[5.4. Exploration Dataset Analytics 8](#_Toc161317452)

[VI. ANALYSIS AND RECOMMENDATION 8](#_Toc161317453)

[6.1. Job trending 8](#_Toc161317454)

[6.2. Key insight and recommendation 8](#_Toc161317455)

[VII. CONCLUSION 8](#_Toc161317456)

[7.1. Difficulties 8](#_Toc161317457)

[7.2. Future Work 8](#_Toc161317458)

[7.3. Perspective 8](#_Toc161317459)

[REFERENCES 1](#_Toc161317460)

[Appendices 1](#_Toc161317461)

[Report Format 3](#_Toc161317462)

# LIST OF FIGURES

# LIST OF TABLES

**Note:** Using **auto generate** Table in Microsoft Word.

# LIST OF ABBREVIATIONS

**An abbreviation** is a shortened version of a term or phrase, e.g. kg for kilogram or Dr. for doctor.

.

# INTRODUCTION

## Presentation of Internship

At the Cambodia Academy of Digital Technology (CADT), My first internship provided a valuable opportunity to apply my knowledge and develop both hard and soft skills in a real-world setting. At the Institute of Digital Research and Innovation (IDRI) located within CADT's innovation center, I spent three months as a data analyst working on the Cambodia Job Trends and Insights (2021-2023) project. This project focused on creating data visualizations of job trends, providing insight that can be valuable to use by both internal and external stakeholders to inform job searches, decision-making, and skill development efforts.

### Objective of Internship

The purpose of CADT’s internship program is for students to apply their theoretical knowledge to real-world applications relevant to their field of study during the internship for three months. The internship program also provides opportunities for students to learn something new and gain more ideas to enhance their ability and energy in the workplace for their next internship or future career. Furthermore, they can improve both skills such as hard and soft skills, gain a deep understanding of workflow and talent development, and team collaboration. Students can actively contribute in part of the project development by utilizing their knowledge in preparing reports or documentation culminating in a final presentation that showcases their learnings after the internship ends.

### Duration of Internship

The duration of the internship at Institute of Digital Research and Innovation (IDRI) is stated for Three months that start from 17th January 2024 to 17th April 2024 with working schedule from Monday to Friday between 08:00am to 05: 00pm.

## Presentation of Organization

### General Information of Company

As a public research institution, the Institute of Digital Research and Innovation (IDRI) under the Cambodia Academy of Digital Technology (CADT) provides digital solutions or values addressing the needs of both public and private sectors to contribute to the development of the digital economy and society of the kingdom. In addition, IDRI also fosters entrepreneurship, research development in digital technology, and innovation including entrepreneurship promotion. After that, build a quality research team by enhancing a good working environment, and collaborate with all related stakeholders both local and international for knowledge with technology transfer.

### Services of Company

The following institutes operate under CADT:

* **The Institute of Digital Technology (IDT)** A public higher education institution specialized in digital technology, providing excellent training programs to foster digital talents indispensable for the development of digital economy and society.
* **The Institute of Digital Governance (IDG)** is a driving force for the development of digital governance in Cambodia through training of digital skills for civil servants and advancing the digital transformation in government agencies.
* **The Institute of Digital Research and Innovation (IDRI)** is the first public research institution dedicated to conducting research and development (R&D) in digital technologies, driving innovation, and promoting entrepreneurship in Cambodia.

### Vison & Mission

* Vision

IDRI aim to be Cambodia’s flagship research and innovation institute with international recognition in digital technology that addresses key social and economic challenges amid a rapid digital transformation.

* Mission

IDRI has many missions to achieve:

* Research and development and Technology Transfer on Digital Technology focusing on emerging technologies such as AI, IoT, Data Science, and Cyber Security.
* Promote Digital Innovation and Technopreneurship
* Conduct Rigorous Policy Studies and Research on Digital Technology and its impacts on the economy and society
* The Recognize Digital Talent

### Address & Contact

For further information, information providing below refer to the contact details of CADT:

* Email: [pr@cadt.edu.kh](mailto:pr@cadt.edu.kh)
* Website: [cadt.edu.kh](https://l.facebook.com/l.php?u=https%3A%2F%2Fwww.cadt.edu.kh%2F%3Ffbclid%3DIwAR2qnoqNFx27o6XgRw5m76njS267PTm-R8J7xIO9wNSEwWo2CE1vx_9BGCQ&h=AT2Ckm23S2a65XaQAs7d6dUULecMXf6ORpBORDDyUsB79Kc3Im3rrk7L4o5UXws1TFssABvQGTxc9_A0lwKF_whwYBOsRVKknoA-NfAC002ZEGQfc8jl8JhHSYMebMwlbRpqjSdabbc)
* Instagram: [cadtofficial](https://www.instagram.com/cadtofficial?utm_source=ig_web_button_share_sheet&igsh=ZDNlZDc0MzIxNw==)
* Phone Number: 015 335 877
* Address: Bridge 2, National Road 6A, Sangkat Prek Leap, Khan Chroy

Changva, Phnom Penh

# PRESENTATION OF THE PROJECT

## General Presentation of Project

In response to job market growth, and Cambodia's rapid economic development, the job market is undergoing significant transformations. Cambodia Job Trending and Insights from 2021 to 2023 project aims to advantage of data analytics to formulate these emerging trends. By using Python and its libraries, some key points are considered. Identification of trending jobs to analysis that will reveal which professions are experiencing the highest growth and demand within the job market. Uncovering insightful jobs is also to identify patterns within the data. and the project will aim to pinpoint promising career paths based on the collected information. this project guides job seekers by providing insights gleaned from the data that will equip individuals seeking employment with the knowledge necessary to prepare themselves for success in the evolving job market. with Inform employers this project also intends to benefit businesses by keeping them abreast of the latest job market trends, enabling them to make well-informed decisions regarding recruitment and adaptation to the changing landscape. In addition, emphasizes key considerations to underscore the importance of factors such as qualifications, experience, language requirement, and job level for both job seekers and employers navigating the dynamic job market to shape their potential.

## Problematic

Choosing a major for studying, looking for a job or internship, skill development, or gaining more potential to adopt future jobs are identified as impact issues for fresh graduates from high school or university. In addition, some challenges of job vacancy posting analysis are addressed based on diverse structures across different sources, the dataset is collected from various sources or platforms such as the HR website, etc. However, the project was launched to deliver valuable insight for the users and students who deal with the struggle to find their dream job and navigate job demand.

## Objective

This project served on Job analytics based on data collected to be guidelines for jobseekers that help them look for prioritize learning and skill development efforts and deciding on a career life. in this processing, python was used in exploratory data analytics to uncover trends and insights within the job demand through data analysis based on some points such as identifying in-demand skills and experience, analyzing job growth across industries, understanding salary trends in different sectors, discovering emerging job titles in the job market with creating a professional report for analytics outcome.

* 1. **Planning of Project**

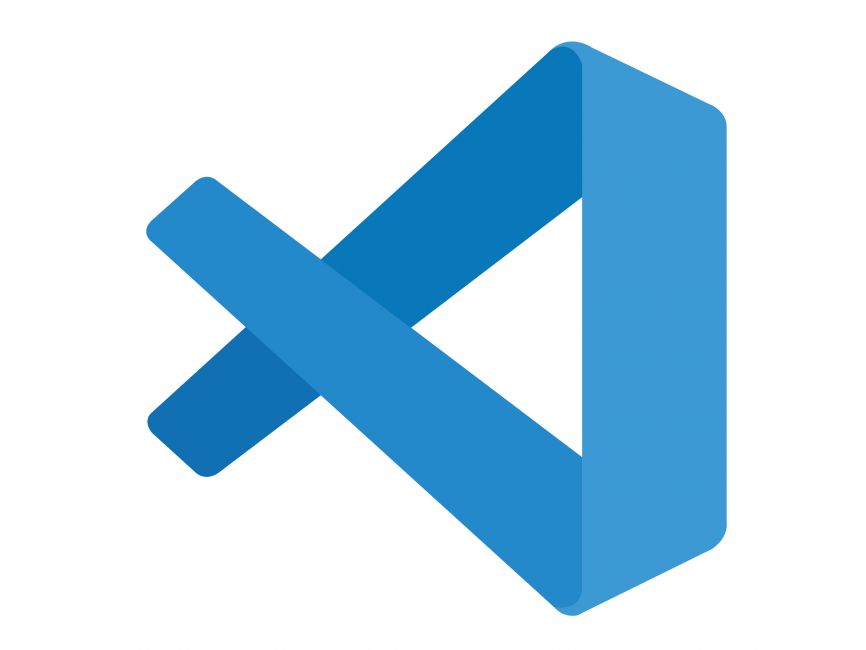
This project took 12 weeks to execute the process and deliver the outcome as our expectation along with our plan and task assigned.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Task | 12 weeks | | | | | | | | | | | |
| W1 | W2 | W3 | W4 | W5 | W6 | W7 | W8 | W9 | W10 | W11 | W12 |
| Research previous report paper |  |  |  |  |  |  |  |  |  |  |  |  |
| Define problem |  |  |  |  |  |  |  |  |  |  |  |  |
| Data Cleaning and processing |  |  |  |  |  |  |  |  |  |  |  |  |
| Data visualization |  |  |  |  |  |  |  |  |  |  |  |  |
| Report and research paper |  |  |  |  |  |  |  |  |  |  |  |  |

# LIERATURE REVIEW

# PROJECT ANALYSIS AND CONCEPTS

## Tools

**Visual Code** is a free open-source code editor developed by Microsoft and available for a variety of operations such as Windows, Linux, and macOS. In addition, it supports the ecosystem of extensions for other programming languages. Some features such as support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git are provided by visual code to serve coding purposes.

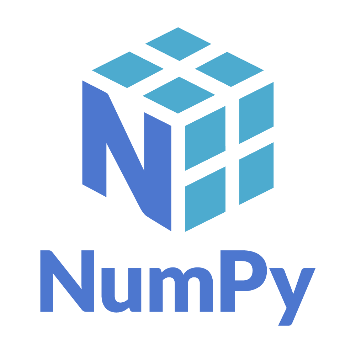


**GitHub** is a platform where developers can create, store, manage, and share code with security and safety. It allows developers to collaborate with their team to push their work efficiency with GitHub’s production.

## Technology requirements

 **Python** is an inventiveness that makes it to be flexible and popular programming language that serves multiple faces such as web development, Machine Learning, software development, mathematics, and system scripting. Furthermore, it is cross-platform unity with simple syntax similar to English. Python runs on a scholar system, meaning that code can be executed immediately after writing. it can treat procedurally, in an object-oriented way, or in a functional way. with the rich ecosystem of free libraries and extensive online support materials, python can adapt to handling large data. its scalability and high across programming languages offer benefits to users. Python is readable and easy to collaborate with other analysts, especially in such ways of communication or maintaining with technical stakeholders when integrated with new data sources and offers solutions for compatibility during handling data step. for data analytics, it has tools to easily identify patterns and explore insight information for a better future.

**Pandas** is a powerful and flexible software library that is written in Python programming language and focuses on managing tools and analysis. it is also a free software library that is released under the three-clause BSD license to build and offer data structures and operations for handling numerical tables and chronological order or tabular data. Besides that, it also highlights its capacities such as data framing, reading and writing data, data alignment, missing data, adjusting, dividing, and fancy indexing. Additionally, it also provides some functions for size mutability, group by, consolidation, hierarchical axis indexing, time series, optimized for performance, academic and commercial.

**NumPy** provides python library provide calculator tools to help you to solve mathematical problems by advanced statistical and mathematical functions. Furthermore, it also works with arrays as well by applying operation into array, making coding shorter and performing matrix operation easily that is centralized for machine learning algorithms.

**Plotly** is open source that provide many interactive charting by plotly.js to serve various purpose such as scientific charts, 3D graphs, statistical charts, SVG geography chart, financial graph, and more.

**Mathplotlib** is also help in data visualisation process by offering graph and management graph with many styles. it has a plotting library and offer an object-oriented API for installing plots into applications using general-purpose.



**Seaborn** is a library python based on matplotlib that offers engaging graph and informative statistical graphics with a higher interface. With minimal code, it creates plot or graph and visualizes random distributions.

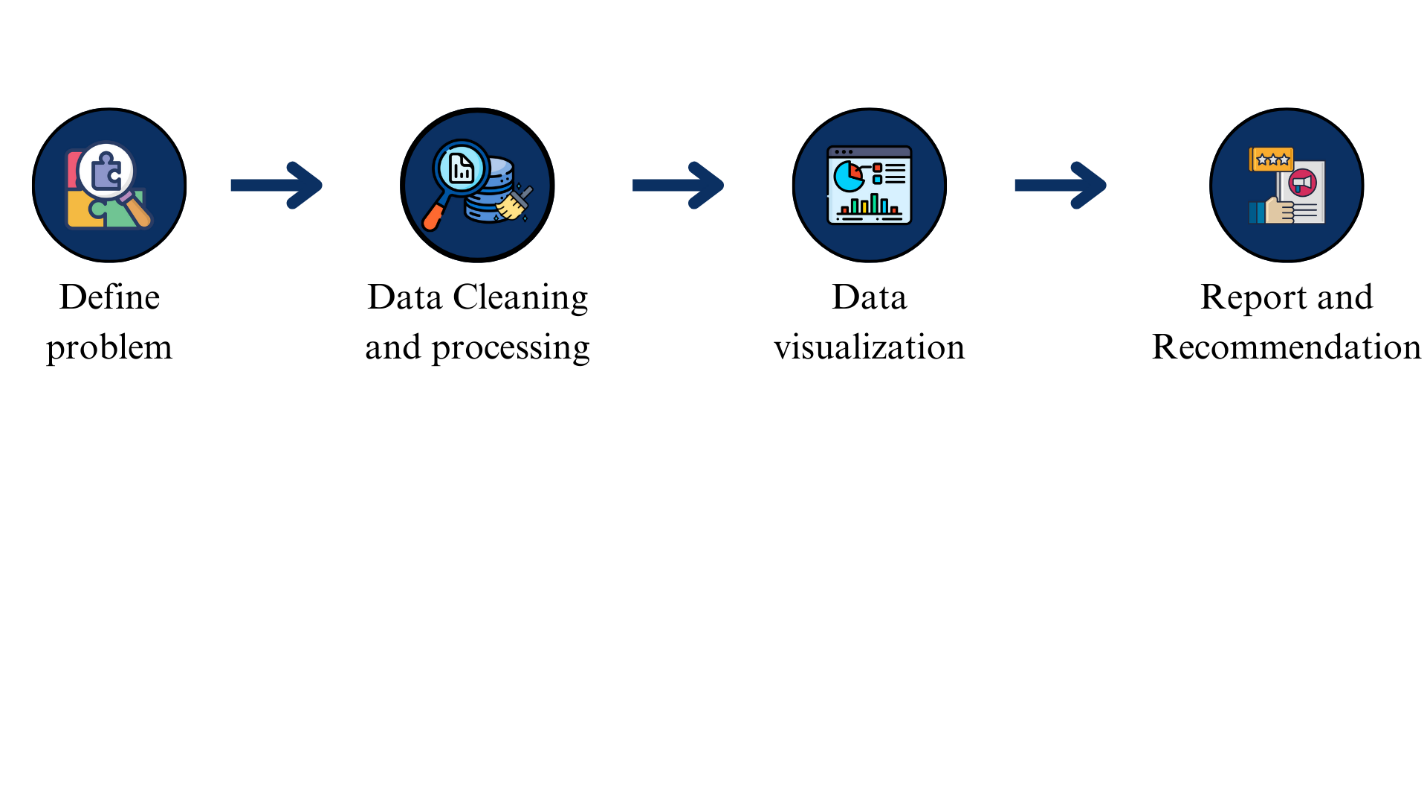


**SciPy** is an open and free source of python library that is used as data-processing library in scientific and technical computing. it also serve multipulate such as numerical integration, interpolation, optimization, , and statistics.



# METHODOLOGY

## Project Pipeline



**Define problem:** we define some key metrics to get insight and look into the pattern entire dataset. by following this step we have listed them below:

|  |  |  |  |
| --- | --- | --- | --- |
| No | Metrics category | Specific metrics | Objective |
| 1 | Job position | * Number of job postings per industry * Percentage change in job postings for specific titles | * Understand industry hiring trends. * Identify growing job sectors. Identify the most frequently appearing job positions * Understand the current job market demand and the types of roles that are in high demand. |
| 2 | Qualifications | * Percentage of jobs requiring specific educational degrees * Percentage of jobs requiring specific years of experience | * Discover in-demand skills within Cambodia. Analyze trends in educational requirements. * Understand experience expectations for various roles. * Explore the most commonly required qualifications ('Qualification' column) Provide insights into the skills and qualifications in demand within the job market. |
| 3 | Salary Trends | * Top salary for each industry * Analyze the salary range ('Salary' column) based on different job positions, industries, or job categories ('Position', 'Industry', or 'Job\_Categories' columns). | * Track salary changes across different sectors. * identify the industries or positions that offer higher salaries and provide insights into the relationship between job type and compensation. |
| 4 | Year of Experience | * Explore the distribution of years of experience ('Year of Exp' column) across different job positions, industries, or levels ('Position', 'Industry', or 'Level' columns). | * to determine the typical experience level expected for different roles and industries. |
| 5 | Gender | * Investigate the gender distribution ('Sex' column) across job positions, industries, or job types ('Position', 'Industry', or 'Job\_Type' columns). | * This analysis can provide insights into gender representation in different sectors and job roles. |
| 6 | Age | * Analyze the age distribution ('Age' column) for different job positions or industries ('Position' or 'Industry' columns) | * to identify any age preferences or trends in specific sectors. |
| 7 | Hiring | * Examine the hiring patterns ('Hiring' column) over time ('Publish Date' column) to identify any seasonal or yearly trends in job postings. | * To understand the fluctuations in job opportunities throughout the specified period. |
| 8 | Location | * Analyze the distribution of job postings by location ('Location' column) to identify the regions or cities with the highest job opportunities. | * to help job seekers target specific locations for better employment prospects. |
| 9 | Language | * Percentage of Top Languages that are needed for job positions. Percentage of language within Top 10 Job Categories and industries | * Explore the most commonly required preferred languages ('Language' column) for different job positions or industries. |
| 10 | Job type | * Percentage of top Job types | * led job seekers to know about what type of job that industries need the most. and help them to well prepare for career life. |
| 11 | Emerging Trending | * Number of hiring in each top industry and job categories appearing in 2023 compared to 2022 and 2021 | * Identify new job categories within industries.  Understand trends in work flexibility. |

**Data cleaning and processing**: rely on the dataset that is collected by scraping from the HR job website in Cambodia, data cleaning and preprocessing have been processed to ensure the dataset can be quality and integrity on some points such as checking missing values, handling duplicating data, detecting English words, and cleaning each column, especially position columns is the most consideration. after that, we can ensure our data is in the right format, consistency, and compatibility for the data visualization.

**Data visualization:** By cleaning of job posting dataset, we can gain more insight and identify the trending jobs in Cambodia between 2021 to 2023. In addition, we also can determine factors that can make job posting data change year by year. By exploring data, We also can highlight the signification and key job market metrics and trends.

**Report and recommendation:** provide recommendations for stakeholders such as employers, educators, and job seekers based on the analysis and insights from visualization.

## Overview Data

With data collection, 94626 rows and 23 columns recorded of the job posting were collected by saving as a CSV file. The dataset was scraped from Cambodia's HR job website and had 18 feathers for the data visualization step.

1. Job Url: Link to the job details on the HR job website.
2. Job title: Name or designation given to a job or position.
3. Company Name: Name of the company with enterprise size.
4. Position: Description of day-to-day responsibilities and tasks in a company.
5. Level: Classification of power and responsibility within a company.
6. Year of Exp: Broad classification of a person's ability based on experience.
7. Hiring: Number of people being hired for specific positions.
8. Salary: Monthly allowance earned by the employee.
9. Sex: Employee's gender.
10. Age: Employee's age.
11. Term: Type of job, such as full-time or part-time.
12. Function/Category: Categorization of job listings for relevant searches.
13. Industry: Group of occupations producing a specific product or service.
14. Qualification: Education level required for the job.
15. Language: Language requirement for the job.
16. Location: Place where the job is performed.
17. Job Description: Outline of duties and responsibilities for the role.
18. Job Requirement: Specific qualifications or skills needed for the job.
19. Company Profile: Background information about the company.
20. Publish Date: Date of job application publishing.
21. Closing Date: The date when the job application closes.
22. Contact Info: Information on how to communicate for the job.

## Data preprocessing

## Exploration Dataset Analytics

# ANALYSIS AND RECOMMENDATION

## Job trending

## Key insight and recommendation

# CONCLUSION

## Difficulties

Technical and communication skill are difficulty that I face during my first internship. In data preprocessing step, I need to understand or identify specific pattern data and choose the right technical for applying to solve these issues. Moreover, I must do research on google and ask a favor from seniors at workplace to deal with this obstacle. Lack of experience initially, I must be flexibility and ingenious in way interpreting my task, algorithm in cleaning dataset and collaborating to find solution from feedback my supervisor, advisor, and seniors to accommodate in environment workplace.

## Future Work

Based on our data, the Jongdoeng Job website is the project's next step. Jongdoeng Job is a dashboard Job trending report web application that is a future to benefit stakeholders. Job hunters, companies, and universities are target users of this platform. Along with this platform, it aims to provide reports, Job Search Integration, Alerts & Notifications for users and keep updates about job vacancies. Moreover, users can use features such as Location Focus, Salary Insights, and Interactive Features to review or detail relevant job information with a friendly user interface.

## Perspective

When the internship started, I was really interested in learning about data preprocessing and visualization methods used in data analytics. I was surprised when I did get deep down into that area and find something new like identify pattern data, making questions to get insight, and use graph in right ways. I also face one of the challenges in this project is detect English word relevant Job title in data preprocessing step. However, by receiving guidance from my supervisor and seniors in the workplace, I can understand data cleaning technique and transformation data. During the internship, I also learned both things soft and hard skills. I am grateful for the opportunity to develop myself by how to communication with advisor, supervisor also other employees in workplace. Additionally, I can improve my problem-solving skills and python coding skill especially new libraries such as PlotLy, WordCloud, and SciPy that is used to execute in this project.

# REFERENCES

**A reference** is a referral to another text or information and can refer to written sources (books, reports, articles).

**Note:** References should be written or generated in website by following APA citation style. (APA7 is recommended).

**Website:** https://www.citefast.com/?s=APA7#\_Webpage

## Appendices

Thank you!

# Report Format

**Margin, Line Space, Font Size**

