Project Title: Study on Job Market Trends & Demand 2023~2024

1. Project Background

The job market is constantly evolving, with varying demand across sectors, roles, and experience levels. Employers post thousands of positions daily, each with different compensation, experience requirements, employment types and posting lifecycles. Analyzing these job postings provides valuable insights into high-demand roles, competitive salaries, and hiring trends. This project examines a dataset of over **1 million job postings** between Feb 2023 and May 2024, to understand market demand, identify emerging opportunities, and highlight patterns in job availability, application engagement, and compensation. The findings aim to support job seekers, recruiters, and policymakers in making informed career, recruitment, and workforce decisions.

2. Project Objective

A. **Analyze market demand:** Identify high-demand job categories and roles

B. **Salary benchmarking:** Examine average and median salaries across sectors, job tiles and position level

C. **Understand hiring patterns:** Explore posting lifecycles, repost frequency, application trends and conversion rate

D. **Support job seekers:** Provide insights for career planning, target roles, and expectations

E. **Inform recruiters & policymakers:** Highlight sectors with skill shortages or high competition

3. Target Persona

👩‍🎓 **Persona 1: Fresh Graduates**

- **Audience:** University and polytechnic graduates
- **Role:** Seeking entry-level roles with defined experience requirements
- **Goals:** Identify high-demand roles offering competitive salaries

- **Challenges:**
    - Limited work experience
    - Unfamiliar with industry expectations and salary ranges

💼 Persona 2: Mid-Career Professionals

- **Audience:** Professionals looking to switch sectors or advance their careers
- **Role:** Seeking roles aligned with skills and salary expectations
- **Goals:** Find high-demand positions and evaluate compensation trends
- **Challenges:** Navigating career transitions and identifying sectors with opportunities

🤝 Persona 3: Recruiters / HR Managers

- **Audience:** HR professionals and talent acquisition teams
- **Role:** Manage recruitment strategy and hiring pipelines
- **Goals:** Understand talent availability, posting competitiveness, and industry benchmarks
- **Challenges:** Sourcing qualified candidates efficiently and optimizing job postings

🎓 Persona 4: Career Advisors / Educators

- **Audience:** Career guidance counselors and educational institutions
- **Role:** Support students or professionals in career planning and skill development
- **Goals:** Align skills with market demand and emerging trends
- **Challenges:** Keeping up-to-date with industry changes and high-demand skill sets

4. Project Specifications

🔧 Tools & Technologies Used

This project utilizes several key Python libraries and tools to perform data cleaning, analysis, and visualization. Each plays a specific role in ensuring efficient data processing, insightful exploration and team work.

🐼 **Pandas** – For loading, cleaning, transforming, and analyzing structured data. Its DataFrame tools make handling large datasets intuitive and efficient.

🔢 **NumPy** – Provides core numerical and mathematical functionality, supporting array operations and calculations that are essential for data preparation and analysis.

📈 **Matplotlib** – A foundational visualization library used to create clear charts and graphs to illustrate job market trends over time.

🎨 **Seaborn** – Built on Matplotlib, this library produces more visually appealing statistical plots, making it easier to explore distributions and correlations in employment data.

💻 **VS Code** – A versatile coding environment for writing, testing, and debugging Python code. Its extensions and Git integration support efficient development workflows.

📊 **NotebookML** – Transform reports from Google Docs into infographics, videos, and audio summaries for interactive and multi-format presentation..

🔗 **Git & GitHub** – Version control and collaboration tools that support teamwork, track changes, and manage project code and documentation in shared repositories.

🤖 **Google Gemini** – Assisted with generating code snippets for data cleaning, transformation, and visualization.

🤖 **ChatGPT** – Assisted with analysis planning, summarizing insights, writing polished project findings and conclusion.

📊 **Streamlit** – **Streamlit** is a **Python-based framework** for creating **interactive web dashboards and data apps** using only Python—no HTML, CSS, or JavaScript needed. It lets Python users quickly build code-based dashboards with **charts, tables, filters, sliders, and buttons**, all displayed in a web browser. Streamlit is commonly used for **data exploration** and **analytics dashboards**.

📊 **Tableau** – A powerful data visualization and business intelligence tool for building interactive dashboards and reports that communicate insights on workforce demand and sector trends.

📊 Dataset Overview – SG Job Data 2023-2024

- **~1.05M job postings** covering roles, categories, and seniority levels
- **Salary data** (min, max, average) for compensation comparison
- **Posting timelines & reposts** to assess hiring activity and urgency
- **Demand indicators**: vacancies, views, and applications
- High data completeness, suitable for **trend analysis and dashboards**

📈 Deliverables

5. Report
6. Interactive Dashboard

🤖 Project Execution / Approaches & TimeLine

This project follows a structured, step-by-step approach to transform raw job data into actionable insights. **MoSCoW prioritization** was used to define key tasks, while a **Kanban board** tracked progress and ensured timely delivery.

Task	Owner	Priority (MoSCoW)	Kanban Status	Duration	Deliverables
Data Collection – gather raw job dataset	Data Analyst	Must Have	To Do	2 days	Raw dataset
Data Cleaning & Transformation – handle missing values, duplicates, outliers, standardize fields	Data Analyst	Must Have	In Progress	5 days	Cleaned dataset
Exploratory Data Analysis (EDA) – summarize distributions, job categories, employment types, salaries	Data Analyst	Must Have	To Do	5 days	EDA report with charts
Correlation & Conversion Analysis – compute correlations, conversion rate, seniority vs salary	Data Analyst	Should Have	To Do	4 days	Analytical insights
Top Job Titles & Salary Analysis – high-volume job titles by category, median salaries	Data Analyst	Should Have	To Do	3 days	Insights tables & charts
Dashboard Design & Development – interactive visuals for categories, positions, salary, applications	BI Developer	Must Have	To Do	5 days	Interactive dashboard (Tableau / Streamlit)
Report Compilation – summarize insights, charts, infographics	Project Lead	Must Have	To Do	3 days	Final report / infographics
Review & Validation – check data, visuals, dashboard functionality	Project Lead & QA	Must Have	To Do	2 days	Validated report & dashboard
Final Delivery & Handover – present to stakeholders	Project Lead	Must Have	To Do	1 day	Delivered project package

5. Project Tasks

Activity 1: Data Loading and Exploration

This activity initializes the analysis by loading and exploring the dataset, ensuring data integrity and format consistency for accurate downstream analysis. After reviewing the dataset in VS Code (data info, head, and basic checks), the following actions were need to performed:

1. **Optimize numeric fields:** Convert average_salary from float64 to int64 to reduce memory usage and improve processing efficiency.
2. **Standardize dates:** Convert metadata_expiryDate, metadata_newPostingDate, and metadata_originalPostingDate to datetime format for accurate time-based analysis.
3. **Transform categories:** Convert the semi-structured categories JSON field into a structured tabular format for easier analysis and reporting.
4. **Handle missing values:** Remove fields and rows with a high proportion of missing data; apply appropriate imputation to columns with low missing rates based on their type and context.
5. **Remove duplicates:** Identify and drop duplicate entries to ensure data quality.

Activity 2: Data Cleaning and Feature Engineering

This activity enhances data relevance and structure, enabling focused analysis on key job indicators and application conversion factors.

After completing the cleaning steps, the dataset contains 1,044,597 entries and 21 columns.

Original JSON-format fields were removed, and the cleaned dataset was saved and exported for use in building interactive dashboards in Streamlit.

Activity 3: Exploratory Data Analysis (EDA)

This activity provides a Job Market Overview, identifying key patterns and anomalies that shaped dashboard design.

Key Insights:

Top Job Categories: Information Technology, Engineering, Administration/Secretarial, Customer Service, and Sales/Retail show sustained hiring demand, highlighting the need for technical workforce supply.

Job Titles: The most frequent titles (“Supervisor”) are broad; analyzing titles together with job categories provides more actionable insights.

Companies: Top 15 companies are predominantly recruitment agencies, indicating their key intermediary role in job postings.

Posting Dates: 4.1% of entries have new posting dates differing from original dates (avg. difference 2.73 days). The new posting date is used for subgroup analysis.

Posting Duration: Average time between posting and expiry is ~26 days (median 30 days), indicating postings remain active for roughly one month.

Posting Volume: Daily postings fluctuate between 500–3,500, with peak monthly postings (>85,000) from June to August 2023.

Data Transformation for Analysis:

- Employment Type: Consolidate Temporary, Internship/Attachment, Freelance, and Flexi-work into “Temporary/Freelance”.
- Minimum Years of Experience: Group 9+ years into “9+”.
- Number of Vacancies: Group 9+ vacancies into “9+”.
- Salary: Values ≥25,000 grouped as “25,000+”, imputed with group median.
- Applications: Values ≥10 grouped as “10+”, imputed with median.
- Views: Values ≥80 grouped as “80+”, imputed with median.

This preprocessing reduces skewness, mitigates outlier impact, and improves interpretability while preserving meaningful differentiation.

Activity 4: Correlation Analysis

Identifying correlations between features is critical for designing interactive elements in the dashboard and understanding key drivers in the job market.

Key Findings:

- Very Strong Relationship:

Total job applications vs. total job views: Spearman r = 0.72 → higher visibility drives more applications.

- Strong Relationships:

Average salary vs. position level: r = 0.67
Average salary vs. minimum years of experience: r = 0.67

- Strong Relationship:

Position level vs. minimum years of experience: r = 0.65 → seniority aligns with experience requirements.

- Moderate Relationships:

Repost count vs. job applications: r = 0.30
Repost count vs. job views: r = 0.31

- Weak Relationships:

Job applications vs. average salary: r = 0.20

Insight:
While salary is largely driven by merit (experience and level), the success of a posting depends on visibility (views).

Robustness Check:
Correlation patterns remain consistent across methods, confirming the relationships are stable and reliable for further analysis and dashboard interactivity.

Activity 5: Segment Analysis – Salary & Conversion Trends

Analyzing average salary trends across position levels, experience bands, and conversion rates uncovers patterns in high-demand roles and job attractiveness.

Key Findings:

- **Salary & Seniority:**
  - Higher position levels correlate with higher average salaries.
  - Senior Management: Median $9,500
  - Fresh/Entry Level: Median $2,550
  - Professional & Manager: Median $6,000

- **Salary & Experience:**
  - Higher experience generally leads to higher pay.
  - 9+ years: Median $10,250
  - Fresh: Median $2,700
  - Milestone effects:
    - 4–5 years → $5,500
    - Jump to 5–6 years → $8,000
    - 7–8 years → $8,500
    - 9+ years → $10,250

    - Market particularly values 4–5 year and 9+ year experience bands.

- **Position Level & Experience:**
    - Required experience aligns with position level, but more experience does not automatically lead to higher roles.
    - Senior Management: Avg 7 years
    - 4 years → Senior Executive/Professional
    - 5 years → Management
    - Skills and knowledge also influence position level.

- **Conversion Rate (Post Efficiency):**
    - Reposted jobs have higher conversion (~8%) than non-reposted (~1.5%), indicating reposting improves post efficiency.
    - Management-level positions have higher conversion rates, attracting more applications relative to views.

Activity 6: In-Depth Job Category & Title Analysis

Conduct an in-depth examination of specific job categories and job titles to understand their relationship with required experience, position levels, and salary variations.

- Any employment type, job category, or job title can be selected for detailed analysis.
- Enables exploration of salary ranges, position levels, and category distributions within selected segments.
- Example:
    - Category: Information Technology (highest-demand)
    - Top 3 Job Titles:
        - Software Engineer → Median Salary: $6,000
        - Software Developer → Median Salary: $7,000
        - Network Developer → Median Salary: $5,250

This approach provides actionable insights for workforce planning, recruitment strategies, and compensation benchmarking.

Activity 7: Time Series Analysis Capturing the daily dynamics of Jobs and monitoring new jobs opened for a comprehensive overview.

Activity 8: Job Market Dynamics – Demand, Salary & Post Efficiency

Develop dashboards to visualize the **interactions between job categories, job titles, and employment types**, highlighting:

- **Demand trends** across categories and roles
- **Salary distributions** by position and experience
- **Post efficiency** using conversion rates

Tools: Streamlit for interactive dashboards and Tableau for advanced visual analytics.

Activities 9: Regress analysis to understand drivers (in-progress)

6. Executive Summary

This project analyzed over **1 million job postings** to uncover key patterns in the Singapore job market, focusing on **demand, salary trends, position levels, experience requirements, and post efficiency**.

- **Data Preparation:** Raw job data was cleaned, structured, and optimized to ensure accuracy and enable robust analysis. Missing values, duplicates, and outliers were handled systematically.

- **Job Market Overview:** Top job categories include **Information Technology, Engineering, Administration/Secretarial, Customer Service, and Sales/Retail**, highlighting sustained demand, particularly in technical fields. Recruitment agencies dominate postings, serving as key intermediaries.

- **Correlation Insights:**

    - **Very strong correlation:** Job views vs. job applications (r = 0.72) – visibility drives applications.
    - **Strong correlation:** Salary with position level (r = 0.67) and years of experience (r = 0.67).
    - **Moderate correlation:** Repost count with applications and views (r ≈ 0.30).
    - Salary is largely merit-driven, whereas posting success is driven by visibility.

- **Salary & Experience Patterns:**

    - Senior Management commands the highest median salary ($9,500), while fresh/entry-level roles earn $2,550.

    - Experience milestones (4–5 years and 9+ years) show significant salary jumps, indicating market value for mid-career and highly experienced professionals.

- **Position Levels:** Align with required experience but are also influenced by skills and knowledge, not just tenure.

- **Conversion Rates (Post Efficiency):** Reposted jobs and Management-level roles attract higher applications relative to views, indicating better post efficiency.

- **In-Depth Analysis:** Within Information Technology, top-demand roles (Software Engineer, Software Developer, Network Developer) show varying median salaries ($5,250–$7,000), highlighting the importance of role-specific insights.

7. Conclusions and Recommendation

❖ **Conclusions**

1. **Market Demand:** Technical and customer-facing roles dominate postings, requiring strategic workforce planning in these areas.

2. **Salary Drivers:** Position level and years of experience are primary determinants of compensation; mid-career (4–5 years) and highly experienced (9+ years) employees are especially valued.

3. **Posting Efficiency:** Visibility and reposting significantly impact application numbers, more so than salary or position alone.

4. **Skills & Knowledge Matter:** Experience alone does not guarantee higher-level roles; competencies and expertise are key factors in senior positions.

5. **Targeted Recruitment Insights:** Top job titles within high-demand categories show significant variation in salary and experience requirements, emphasizing the need for granular analysis.

❖ **Recommendations**

1. **Workforce Planning:** Focus recruitment efforts on high-demand categories such as Information Technology and Engineering, ensuring availability of qualified mid-career

and senior talent.

2. **Salary Benchmarking:** Align compensation strategies with experience and position-level insights to attract and retain talent, particularly in milestone experience bands.

3. **Job Posting Strategy:** Increase visibility through effective job descriptions, targeted postings, and reposting to enhance conversion rates.

4. **Skills Development:** Invest in training and professional development to enable employees to progress into higher-level positions, recognizing that skills are as important as tenure.

5. **Data-Driven Recruitment:** Leverage dashboards and interactive tools to monitor demand, salary trends, and post efficiency, enabling timely and actionable decision-making.

8. Challenges and Limitations

1. Data Quality Issues

- Missing values were found in the same cases across multiple variables. The root cause of these missing values could not be determined due to the absence of documentation.
- Inconsistent formatting in categorical fields (e.g., *Categories* stored in JSON format or multiple categories recorded within a single column) required additional parsing and restructuring.
- Certain fields contained long distinct value listings rather than standardized categories (e.g., company name, job title), making categorization and aggregation more complex.
- Outliers were observed (e.g., values such as "999"), but without a codebook or metadata documentation, their true meaning could not be verified.

These issues required additional data cleaning, transformation, and assumptions, which may introduce minor processing bias into the analysis.

---

2. Salary Data Constraints

- A relatively high percentage of outliers was detected in the average_salary variable.
- Salary ranges were not standardized across job postings (e.g., monthly vs annual salary, inconsistent reporting formats).
- Some postings did not disclose salary information.

Given these constraints, median salary was used as a more robust central tendency measure instead of the mean. However, salary-related findings may still not fully represent the actual market distribution.

---

3. Repost and Engagement Metrics Interpretation

- Repost count does not necessarily indicate new hiring demand, as it may reflect the reactivation of existing listings.
- A strong correlation between total applications and total views ($r = 0.86$) suggests potential multicollinearity, limiting independent interpretation of these variables.

- Engagement metrics may be influenced by platform algorithms, visibility ranking, or paid promotions rather than actual job attractiveness.

---

4. Time-Based Limitations

- The analysis is restricted to the selected time frame (2023–2024).
- Seasonal hiring patterns may distort monthly comparisons and trend interpretation.
- The dataset does not incorporate macroeconomic factors, labor market policies, or external economic events that may influence job postings.

---

5. Dashboard and Visualization Constraints (Tableau)

- Data blending across multiple sources may introduce aggregation-level mismatches or filter inconsistencies.
- Reference lines (e.g., 2023 vs 2024 averages) depend heavily on filter context and level of detail configuration.
- Dashboard filters may not apply globally unless configured correctly (e.g., “Apply to All Using This Data Source”).

---

6. Analytical Assumptions

- The analysis assumes that posted job listings accurately reflect actual hiring demand.
- Unadvertised roles and internal hiring processes are not captured in the dataset.
- Correlation findings do not imply causal relationships.