**LinkedIn Jobs Data Analysis Report – Power BI Dashboard**

**Project Overview:**

This project focuses on a data-driven analysis of job listings scraped from LinkedIn across three diverse regions: **Canada**, **USA**, and **Africa**. The original dataset, collected from kaggle [Link here](https://www.kaggle.com/datasets/cedricaubin/linkedin-data-analyst-jobs-listings?resource=download) , contained key information such as job title, company name, work mode (remote, onsite, or hybrid), location, and salary information.

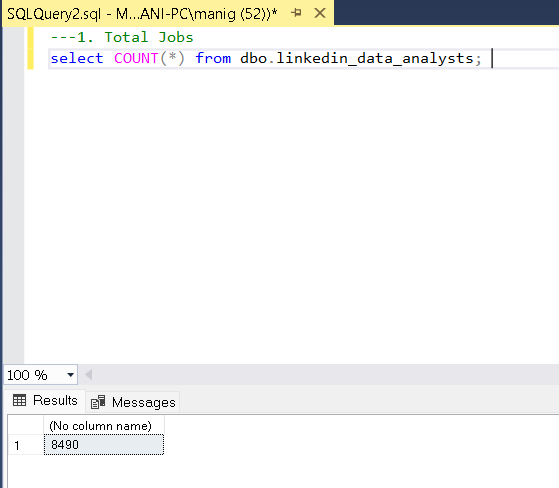
To ensure completeness and high-quality insights, missing salary values were predicted using a robust **XGBoost regression model** within a Python environment in Google Colab, achieving an R² score of **0.9795**, indicating a very strong predictive performance. Once cleaned and enhanced, the enriched dataset was exported and used to build a powerful, interactive dashboard in **Power BI**.

The goal was to empower users with actionable insights into job trends, salary distributions, and company hiring patterns across the three regions.

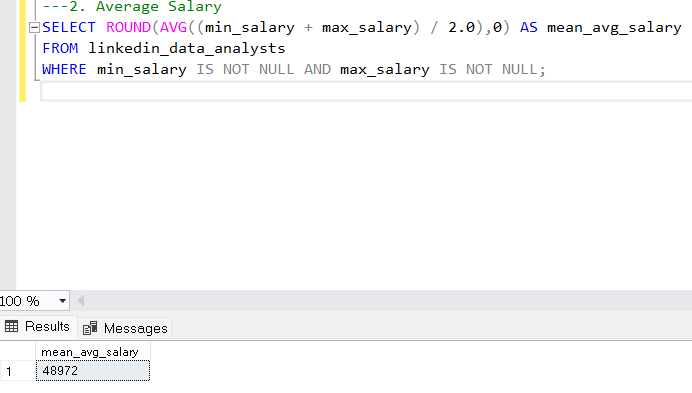
**Inside Power BI Dashboard**

**Key Performance Indicators (KPIs):**

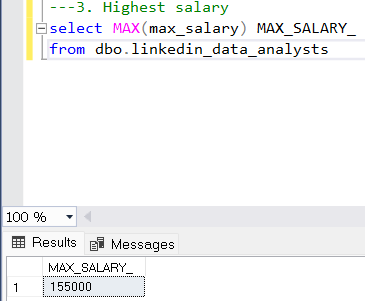
* **Total Jobs**   
  + Definition: Total number of job listings in the dataset.
  + Metric: COUNT(\*)



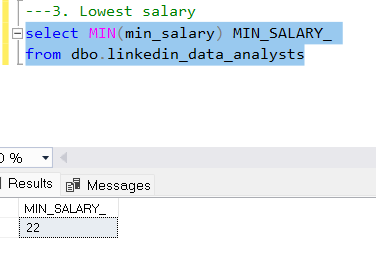
* **Average Salary**
  + Definition: Mean of predicted average salaries across all job listings.
  + Metric: AVG(avg\_salary)



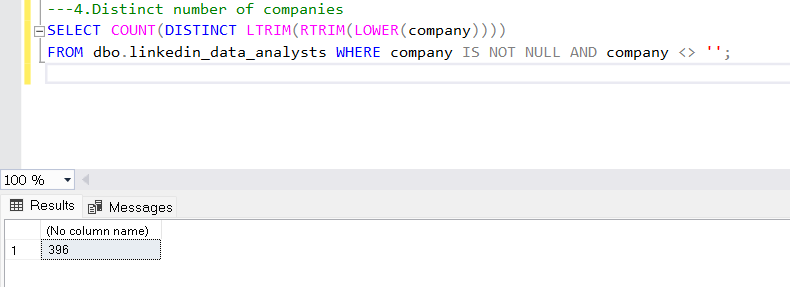
* **Maximum Salary Offered**
  + Definition: The highest salary observed in the entire dataset.
  + Metric: MAX(max\_salary)



* **Minimum Salary Offered**
  + Definition: The lowest salary found in the dataset.
  + Metric: MIN(min\_salary)



* **Total Unique Companies Hiring**
  + Definition: Total number of unique companies posting jobs.
  + Metric: DISTINCTCOUNT(company)

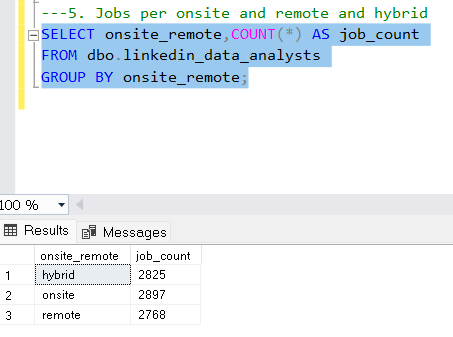


**Interactive Filters & Slicers:**

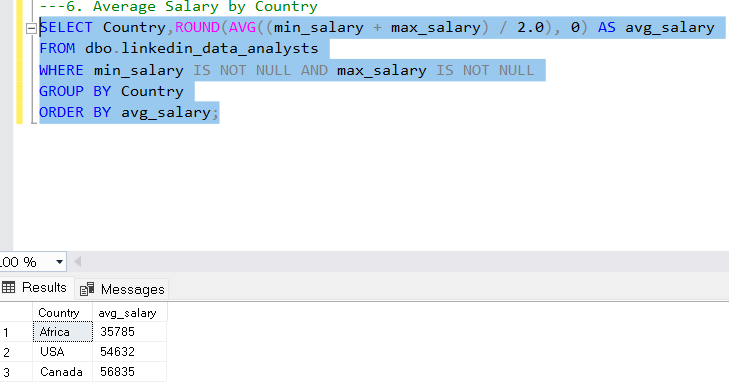
* **Country Selector (Checkbox Slicer)**: Allows users to interactively filter all visuals based on selected countries (Canada, USA, Africa). This interactivity brings regional comparisons to life.

**Core Visualizations and Insights:**

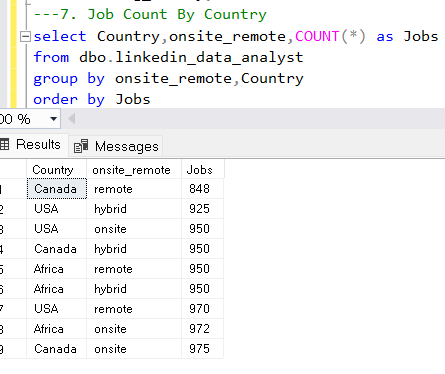
1. **Job Distribution by Work Mode (Doughnut Chart)**
   * Purpose: Helps understand the proportion of job types (Remote, Onsite, Hybrid).
   * Insight: Useful for job seekers interested in flexible work environments.



1. **Average Salary by Country (Bar Chart)**
   * Purpose: Displays how average salaries differ across regions.
   * Insight: Users can identify where high-paying opportunities are more concentrated.

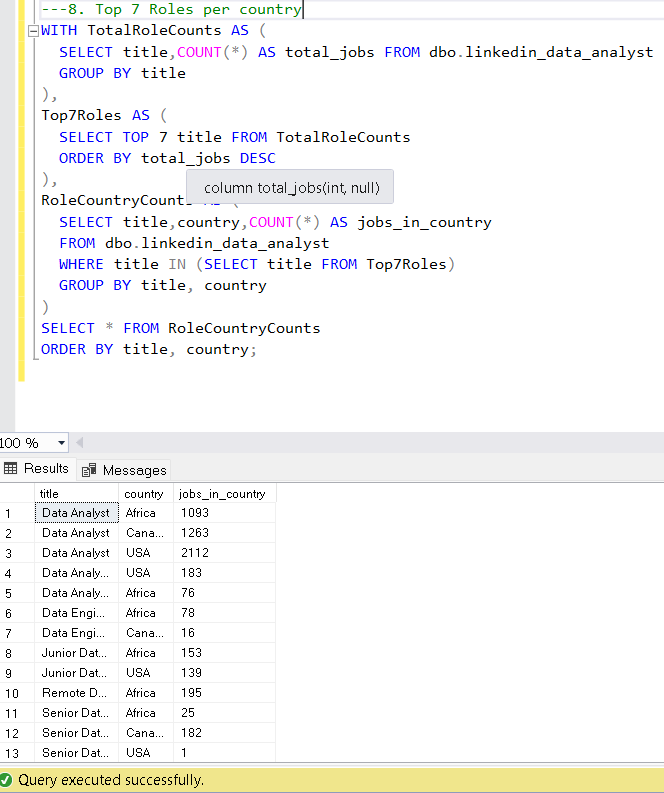


1. **Job Count by Country and Work Type (Stacked Bar Chart)**
   * Purpose: Breaks down job listings by country and further by remote/onsite/hybrid modes.
   * Insight: Useful for comparing hiring strategies and work culture trends regionally.

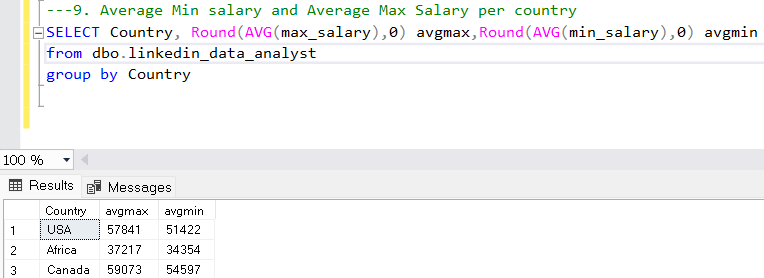


1. **Jobs Over Time (Area Chart)**
   * Purpose: Visualizes job posting frequency over time.
   * Insight: Identifies seasonal or monthly hiring patterns.
2. **Company Salary Summary (Multi-row Card)**
   * Purpose: Shows companies alongside their offered salary ranges.
   * Insight: Highlights top-paying companies and their market competitiveness.

1. **Top 7 Job Roles per Country (Stacked Bar Chart)**
   * Purpose: Visualizes the seven most common job titles per country.
   * Insight: Reveals in-demand roles and title variation by geography.



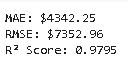
1. **Average Min & Max Salary by Country (Grouped Bar Chart)**
   * Purpose: Compares average lower and upper bounds of salary per country.
   * Insight: Assists in assessing salary fairness and spread in different regions.

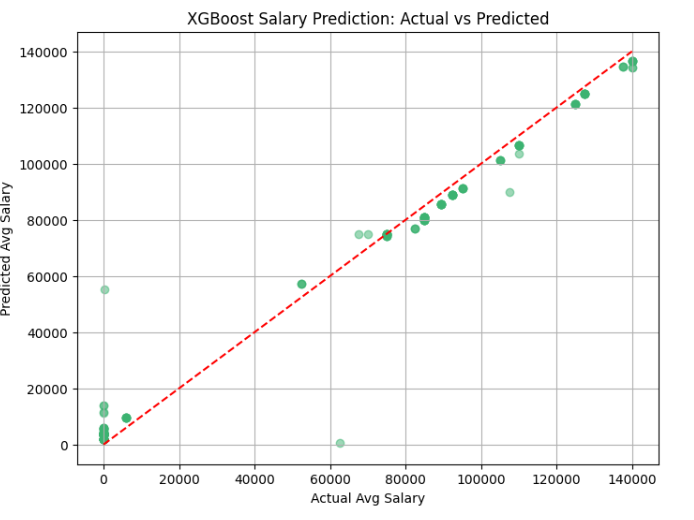


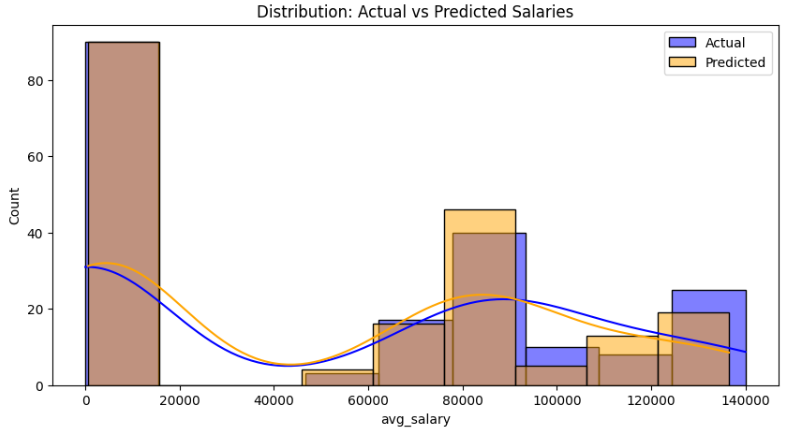
**Advanced Processing Behind the Dashboard:**

* **Salary Imputation via XGBoost Regression:**
  + All missing values in min\_salary and max\_salary were filled using a new column (average\_salary - mean of min and max salary) created using existing min and max salaries .This new column is used to train the XGBoost model and it predicted the values of remaining missing values and gave the missing values in average\_salary. From this, I have separated the min salary and max salary using 96% of average\_salary and 104% of average\_salary.
  + Features used included: job title, company, location, remote/onsite status, and country.
  + Resulting metrics:

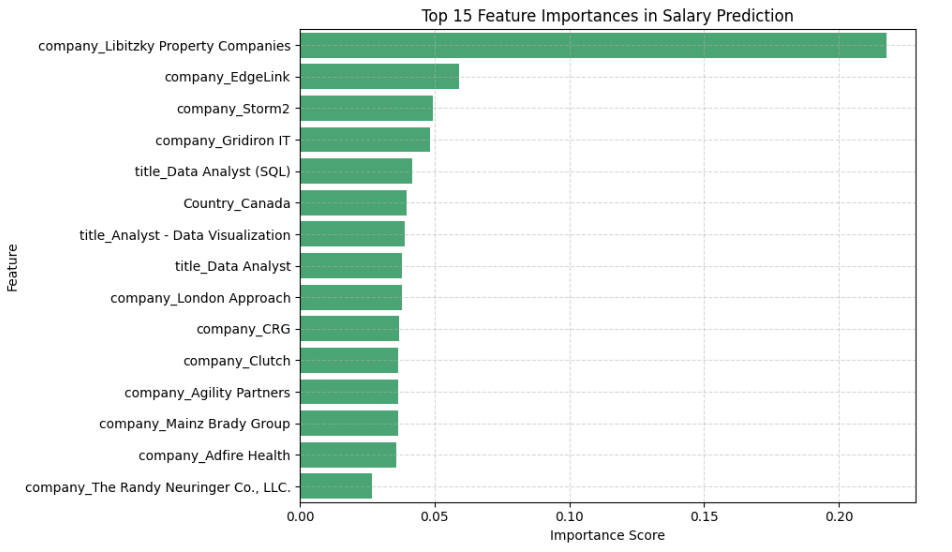
1. MAE = $4,342.25
2. RMSE = $7,352.96
3. R² Score = **0.9795**.



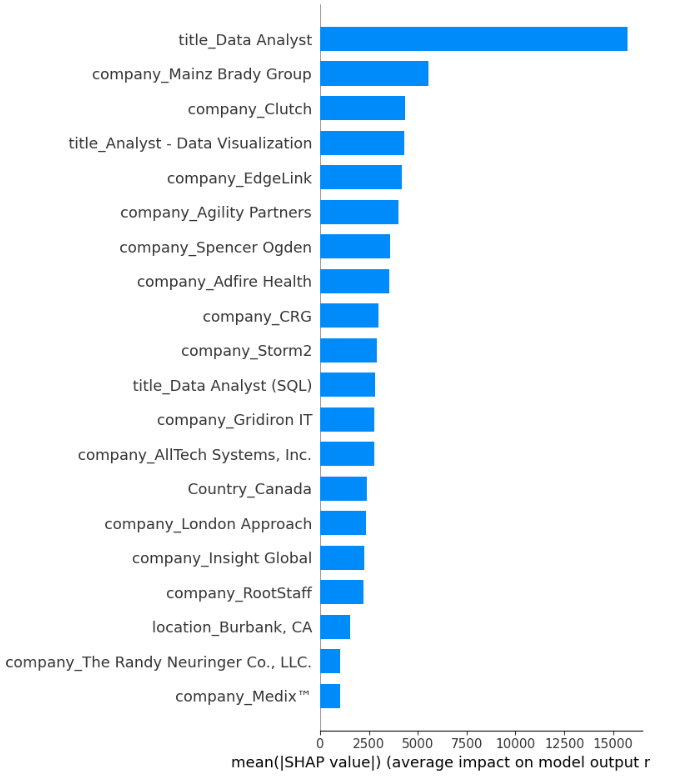




* **Feature Importance Analysis:**
  + Country and location were identified as high-influence features in predicting salaries.
  + The feature importance plot (created in Python) helps interpret the regression model’s decision-making process.



* **Model Explainability using SHAP (Shapley Additive Explanations):**
  + To gain deeper insight into how each feature influenced salary predictions, SHAP values were calculated and visualized.
  + SHAP allows both global and local interpretability, showing which features most consistently impact the model, as well as how they affect specific individual predictions.
  + Visuals such as the SHAP summary plot and beeswarm chart were used to understand non-linear relationships, feature interactions, and the positive/negative impact of features like Country, location, and onsite\_remote.
  + This level of transparency increases trust in the model and highlights how real-world variables contribute to compensation outcomes.



**Conclusion:**

This project successfully integrates data science and business intelligence to deliver a holistic view of the LinkedIn job market across Canada, the USA, and Africa regarding different types of Data Analyst jobs. By cleaning and modeling the data rigorously before importing into Power BI, we ensured accurate and reliable insights.

The final Visualised dashboard is designed for:

* **Job seekers** to explore market demand and top-paying roles
* **Recruiters** to benchmark their offerings
* **Policy makers** to understand employment trends

With interactive slicing, insightful KPIs, and visually rich charts, this dashboard stands as a practical decision-making tool powered by modern data analytics.

**Copyright 2025 Mani Shankar Yalapalle. All rights reserved.**