

## Egyptian Software Market Salary Analytics & Predictive Modeling

2025

Malak Ahmed Saber





## Define Objectives

#### Business Question:

How can we analyze and visualize salary trends in Egypt's tech industry to understand the relationship between experience, job title, work type, and salary?



#### Specific Objectives:

- Analyze salary distribution by years of experience
- Compare salaries across different job titles
- •Examine the effect of work type (Remote, On-site, Hybrid) on salaries
- Identify salary ranges for different cities
- •Build an interactive dashboard for salary insights
- •Build and Deploy ML models to predict salaries, solving the "what should I ask as expected salary?" dilemma

## 2. Identify Stakeholders

#### Stakeholder

#### **Role / Use of Results**

Job Seekers	Understand salary expectations based on experience & role and use the salary prediction tool
Employers / HR Teams	Benchmark salaries against industry averages
Career Advisors	Provide guidance on expected salary ranges
Data Analysts	Derive insights for salary trend reports
Dashboard Users	Explore data interactively for decision-making

### 3. Set Success Metrics

Goal	Metric
Quality of data analysis	At least 5 meaningful salary insights identified
Dashboard usefulness	Filters by role, experience, work type, and city implemented
Data accuracy	100% cleaned and formatted (consistent salary & date formats)
ML model performance	R <sup>2</sup> > 0.8 or MAE < target threshold

## 4. Timeline & Resource Planning

PHASE	ESTIMATED TIME
Data Cataloging	0.5 day
Data Cleaning & Formatting	1–2 days
EDA & Visualization	2 days
ML Model Training	2–3 days
ML Model Deployment	1–2 days
Dashboard Creation	2 days
Documentation & Reporting	1 day



(2)
Initial Data
Cataloging



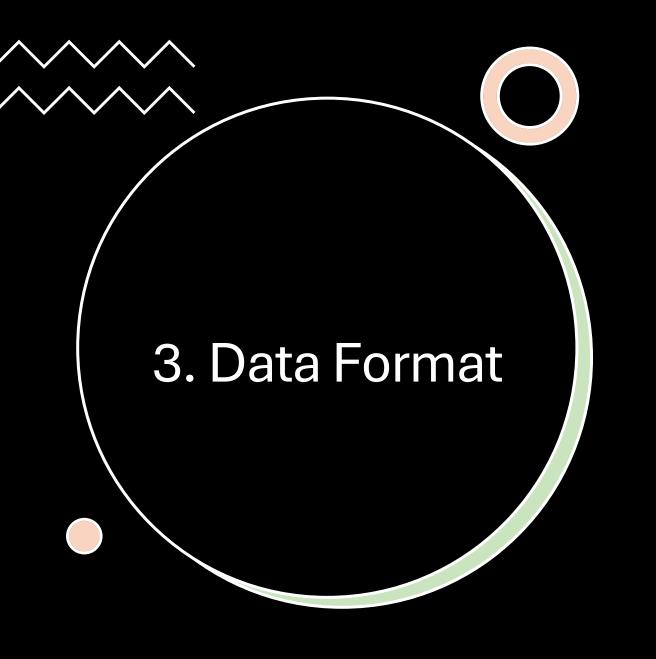
- Name: Egypt Tech Salaries
   2025
- Source : <u>Salaries in Egypt at</u> 2025 ( - (الردود) Google Drive
- Data Collection: Through a survey with google docs



This dataset contains information about job postings and salary offers in Egypt's tech industry.

It captures details such as job titles, years of experience, offered salaries, work type, working hours, company location, and other related attributes.

The data is suitable for **exploratory salary analysis**, **industry benchmarking**, **and predictive modeling** for salary estimation.



•File Type: CSV / Excel (tabular format)

•Number of Rows: ~1,000

•Number of Columns: 9

•Size: Small (< 1 MB)

•Structure: Each row represents a single job posting or salary record.

•Encoding: UTF-8 (to support English and Arabic text)

## 4. Column Summary (Schema)

Column Name	Data Type	Description
TimeStamp	string / datetime	The date and time when the entry was recorded (may contain Arabic AM/PM notation).
Title	string	Job title or role (e.g., Front End, IT Specialist, Full Stack .NET).
Years of Experiences	float / string	Number of years of professional experience required or reported (may include text like "1.5 years").
Salary	string / float	Offered salary (may include currency like "EGP" or be numeric only).
Date of Salary	date	The date the salary information applies to.
What Is your Company	string	Company nationality or description (e.g., "Egyptian", "Not Egyptian but site in Egypt").
Work Type	string	Nature of work location (e.g., Remote, On Site, Hybrid).
Work Hour	string	Working hours type (e.g., Full Time, Part Time).
City of Company site	string	City where the company's office is located (e.g., Cairo, Maadi).



#### •Inconsistent formats:

- Salary column has mixed formats (e.g., 20000, 14000 EGP) that require normalization.
- Years of experience can be numeric or text (e.g., 1.5 years, 0).

#### •Encoding issues:

• Time Stamp values include Arabic AM/PM text (e.g., Ù...) which need cleaning.

#### •Missing values:

• Some records may have blank or null salary, work type, or company nationality fields.

#### •Special cases:

- "Not Egyptian but site in Egypt" should be standardized in the Company field.
- Multiple city names in the same field (e.g., Maadi, Cairo) might need splitting for geographic analysis.



#### Analytical Use Cases:

- Salary Insights: Identify average salaries by job title, experience, and work type.
- Experience vs. Pay Trends: Measure how experience impacts offered salaries.
- Location Analysis: Compare salaries across cities or regions.
- Work Type Comparisons: Analyze salary differences between Remote, On Site, and Hybrid jobs.

#### Machine Learning Use Cases:

- Salary Prediction Model: Predict salaries based on title, experience, work type, and location.
- **Clustering**: Group similar job postings based on salary and required experience.
- **Trend Forecasting**: Predict future salary trends based on historical posting dates.

#### Business Use Cases:

- **Employer Benchmarking**: Companies can compare their offers to market averages.
- Career Planning: Job seekers can set realistic salary expectations.
- Market Reports: HR agencies can generate salary reports for the tech sector.



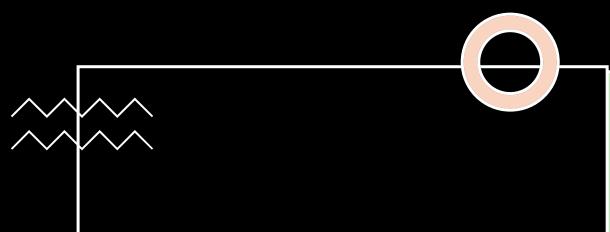
## (3) Data Exploration & Understanding

### **Problems Found**

- No duplicates
- Displayed data.info and found last column "النتيجة" totally null, and city column with 58 nulls
- City column maybe filled according to where the company lies and the salary currency so we fill with the capital of the country
- Columns namings non conventional as data came from a google docs form
- Date column needs to get fixed
- Value counts may need changes for some columns
- Displayed data.describe and columns need data type fixation



# (4) Data Cleaning & Preprocessing



## Steps

- Removed "النتيجة FULLY null column
- Renamed all columns to lower case and conventional names
- Time columns datatype to date time
- Handled years of experience value counts with mapping, then converted data type to int from object
- Same to companyCountry, and changed some values according to worktype
- Same to worktype and work hour and city
- Handled null city according to companyCountry and salary currency
- Made a separate column for currency and converted salary to int
- Converted the 'Title' column to cleaned title by handling case sensitive words, then categorized the jobs into 'category' column according to existence of certain words in the cleaned titile
- Saved the Cleaned Dataset to a csv file for visualization



## (5) Dashboarding

Overview

Salary Analysis

Work & Locations

Currency & Time Trends Average Salary

25.12K

Median Salary

18K

Number of Responses

791

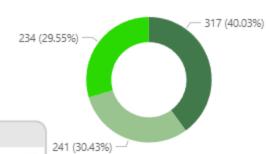
Number of Unique Jobs

443

Number of Unique Cities

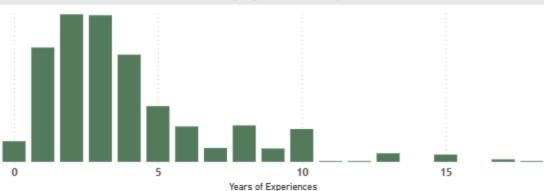
48

#### Work Type Distribution

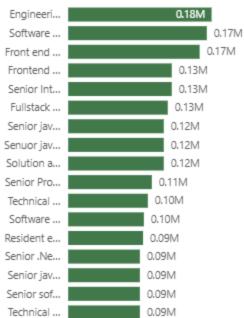


● Hybrid ● Remote ● On Site

Sum of salary by Years of Experiences

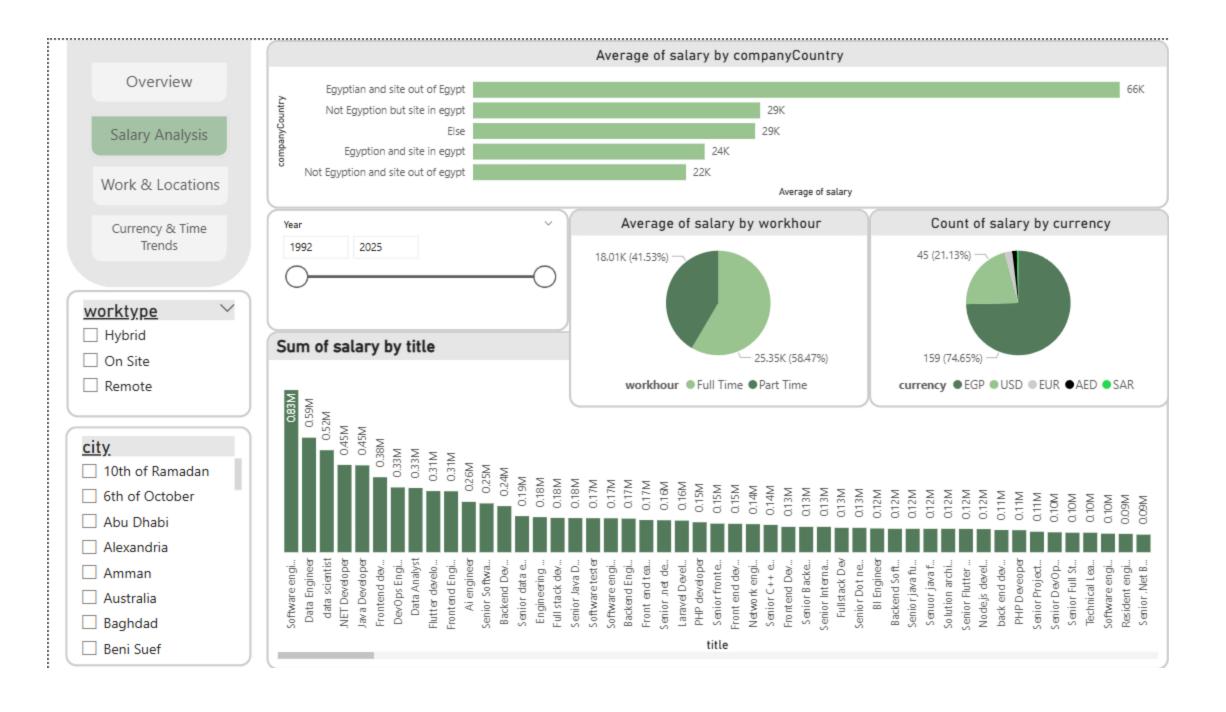


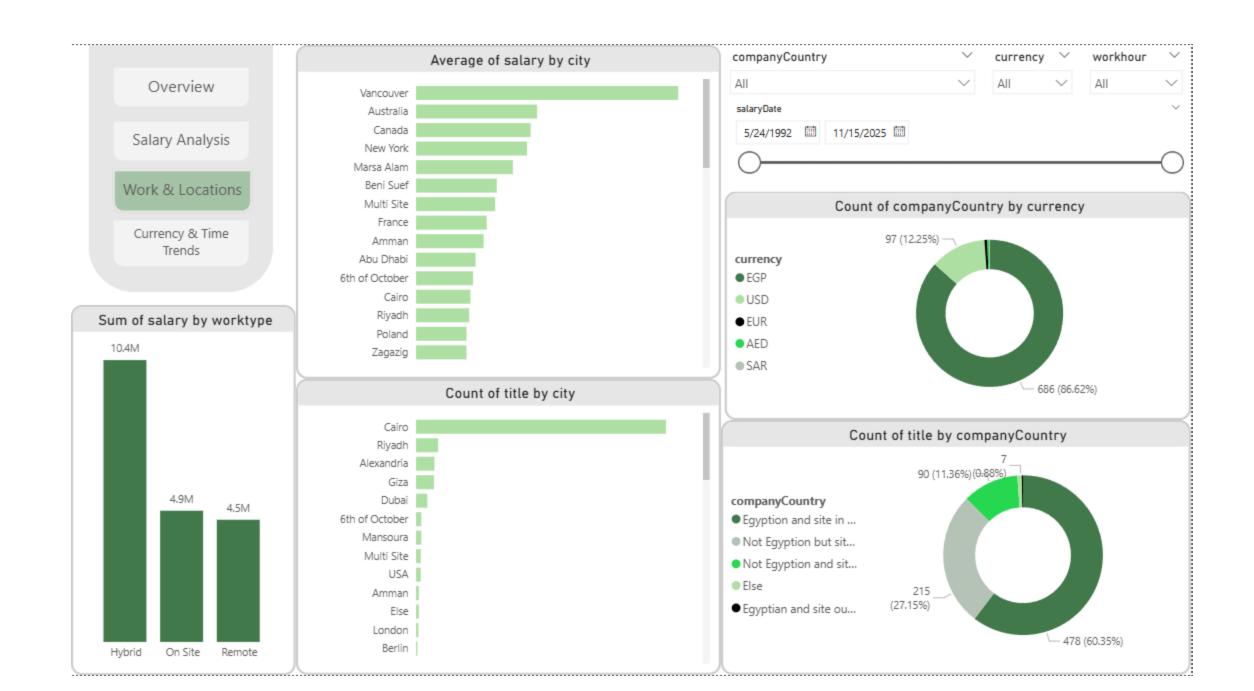
#### Jobs Average Salary

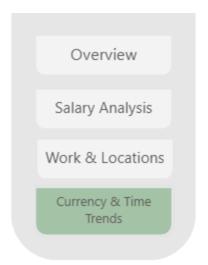


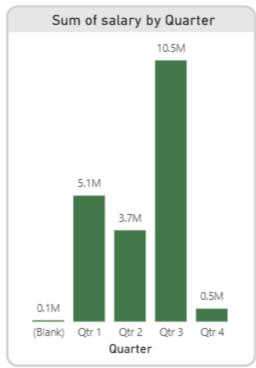
#### Sum of salary across cities and company Location

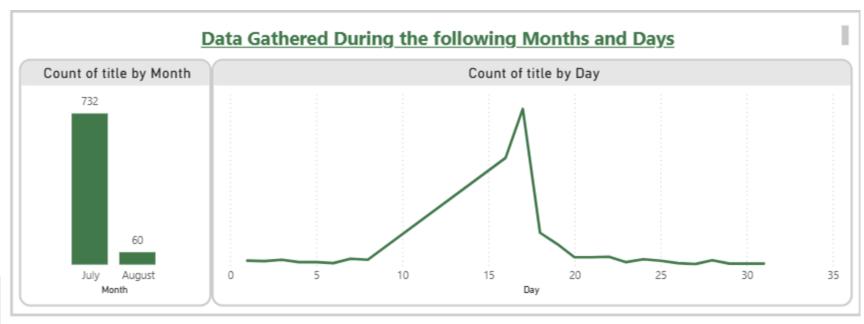


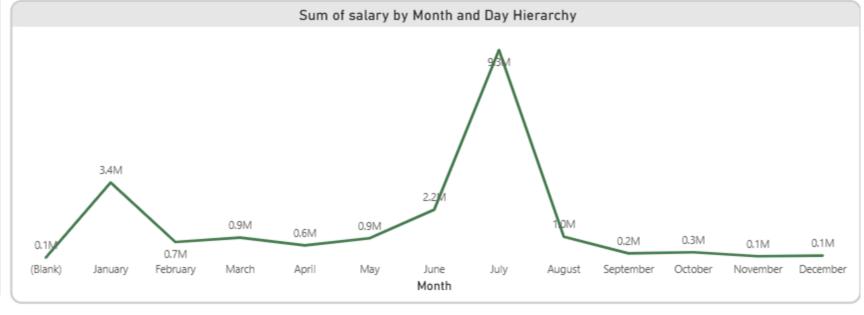














(6) Modelling

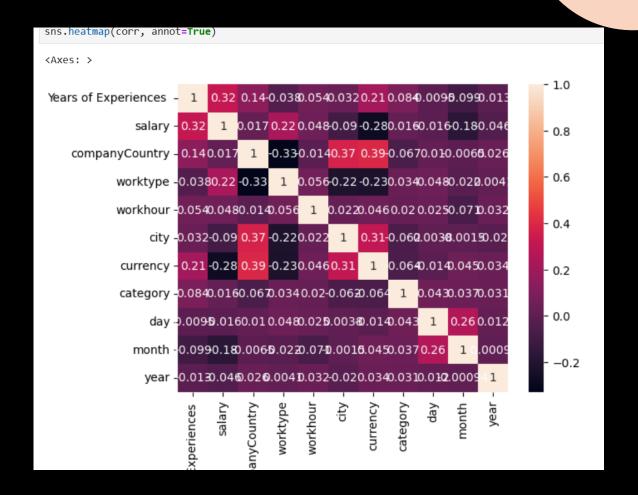


## Steps

- 1.Feature Selection :
  - Divided date column into day, month, year
  - Dropped date, title, timestamp columns
- 2.Encoding:
  - Encoding all categorical columns companyCountry, worktype, workhour, currency, city, category
- 3.Correlation and visualized through seaborn heatmap
- 4.Train Test Split with 'salary' as the target
- 5.Applied Standard Scaler
- 6.Trained Random Forest Regressor
- 7.Evaluated the Model
- 8.Saved the Model using joblib as pickle



## Correlation





# Evaluation of Random Forest

Test Set Evaluation:

MAE: 16135.742231569942

RMSE: 25459.863353250814

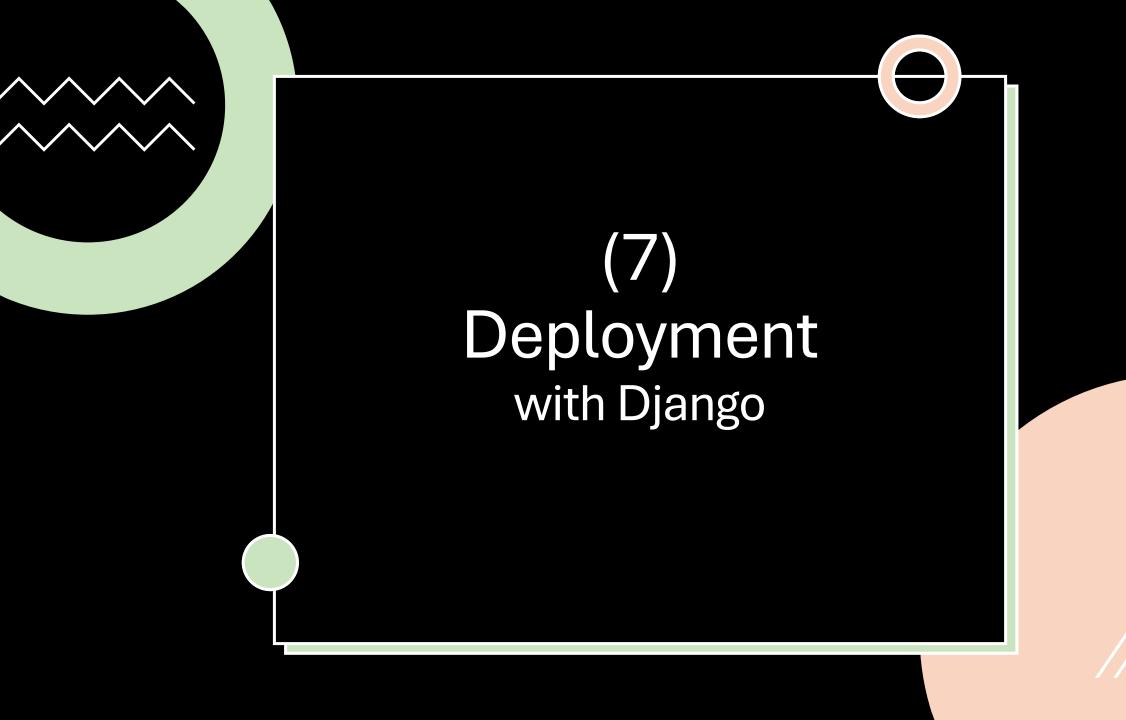
R<sup>2</sup>: 0.31310854978653313

Train Set Evaluation:

MAE: 6203.334565403601

RMSE: 9081.495536222263

R<sup>2</sup>: 0.8746960237883763





Get accurate salary predictions based on your profile

