BUSINESS ANALYSIS

DATA SCIENCE SALARY INSIGHTS

(2020-2024)

Batch

Business Analyst @ masai

Presented by

Anupam Kumar

TABLE OF CONTENTS

SUMMARY OF INSIGHTS

Salary Distribution & Trends

- Salary is right-skewed: Most professionals earn between $75,000–$200,000, with a few high outliers.
- Median salary peaked in 2023 (~$145,000), followed by a slight dip in 2024.
- Overall, strong growth trend seen from 2020 to 2023 (+~80%).

Salary by Company Size

- Medium-sized companies offer the highest median salary (~$145,000).
- Large companies pay moderately ($110,000), while small companies lag behind ($75,000).

Salary by Experience Level

Clear positive correlation:

- Entry-level: ~$75,000
- Mid-level: ~$105,000
- Senior-level: ~$155,000
- Executive-level: >$180,000

Top Paying Job Titles (Median)

- Analytics Engineering Manager: ~$400,000
- Data Science Tech Lead
- Managing Director, Data Science
- AWS Data Architect
- Cloud Data Architect
- High-paying roles are typically leadership or architect-level positions tied to engineering and cloud systems.

Top 10 Paying Countries (Median)

- Israel leads with median salary ~$400,000.
- Followed by Qatar, Malaysia, Puerto Rico, and the US.
- High-paying countries include a mix of developed and emerging economies.

Work Model Comparison

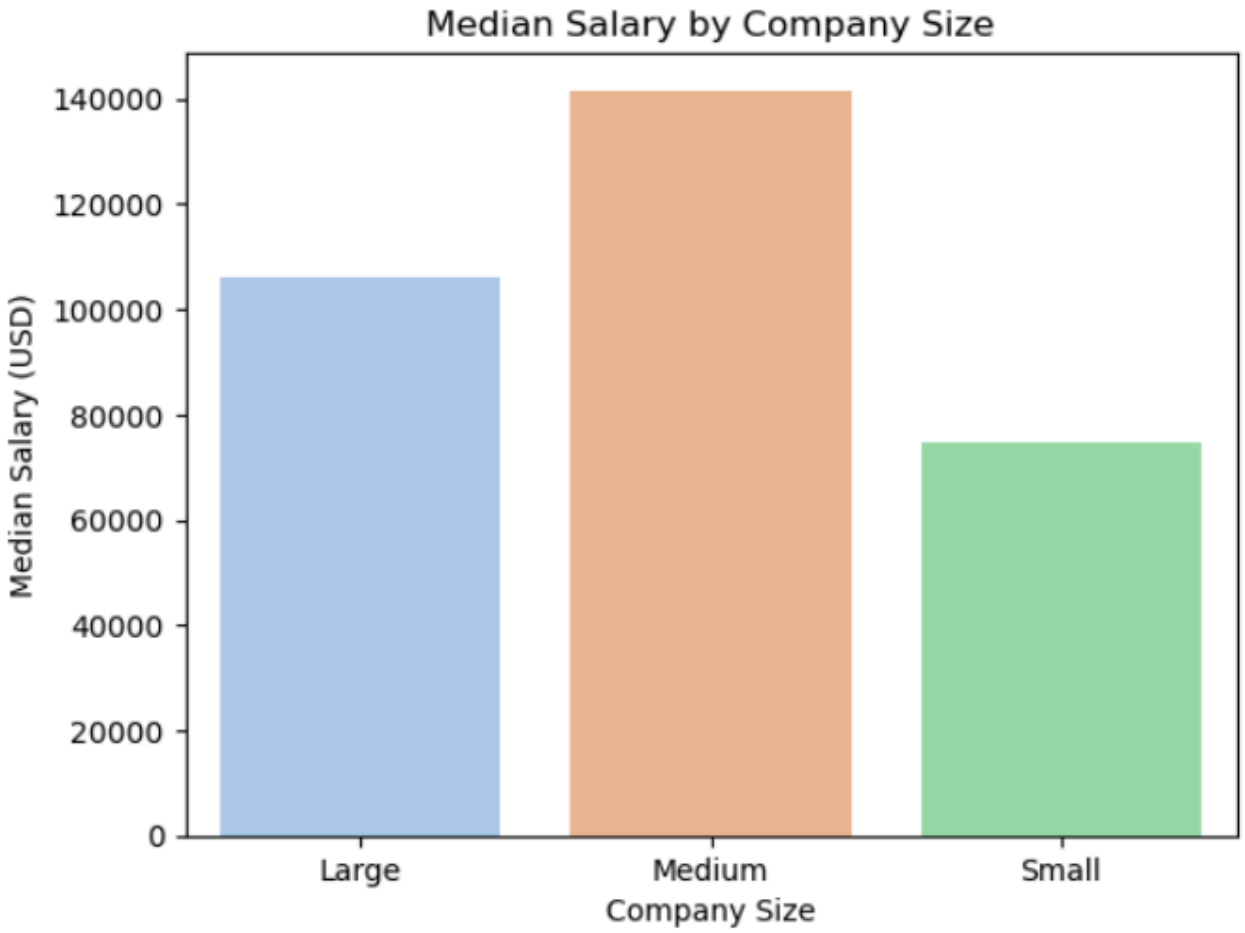
- Remote and On-site roles have comparable median salaries.
- However, remote roles show wider variability with more high-earning outliers.
- Hybrid roles tend to cluster lower in salary distribution.

Correlation Analysis

Strongest positive correlation of salary is with:

- Experience level
- Job title
- Company location
- Work model and employment type show weaker correlation with salary.

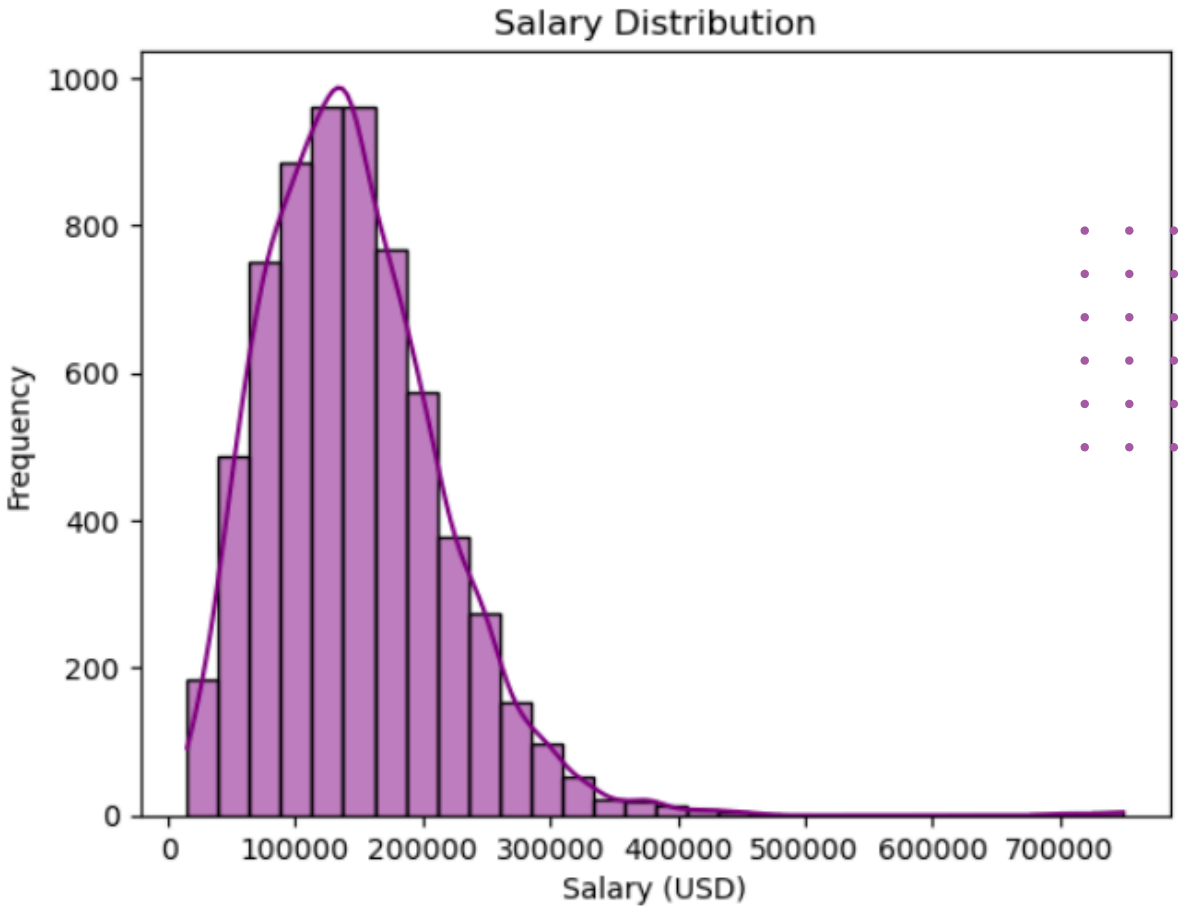
MEDIAN SALARIES BY COMPANY SIZE



This chart shows that medium-sized companies offer the highest median salary, followed by large companies, and small companies offer the lowest.
As a Business Analyst, this insight helps in creating hiring and salary strategies:

1. Small companies should focus on non-monetary benefits like flexible hours, growth opportunities, work-life balance, and strong company culture to attract talent.
2. Medium companies should revise their salary structures regularly to stay competitive and retain top talent.
3. Large companies can continue offering high salaries but must ensure they are getting a good return on investment (ROI) for each data scientist hired, by tracking performance and outcomes.

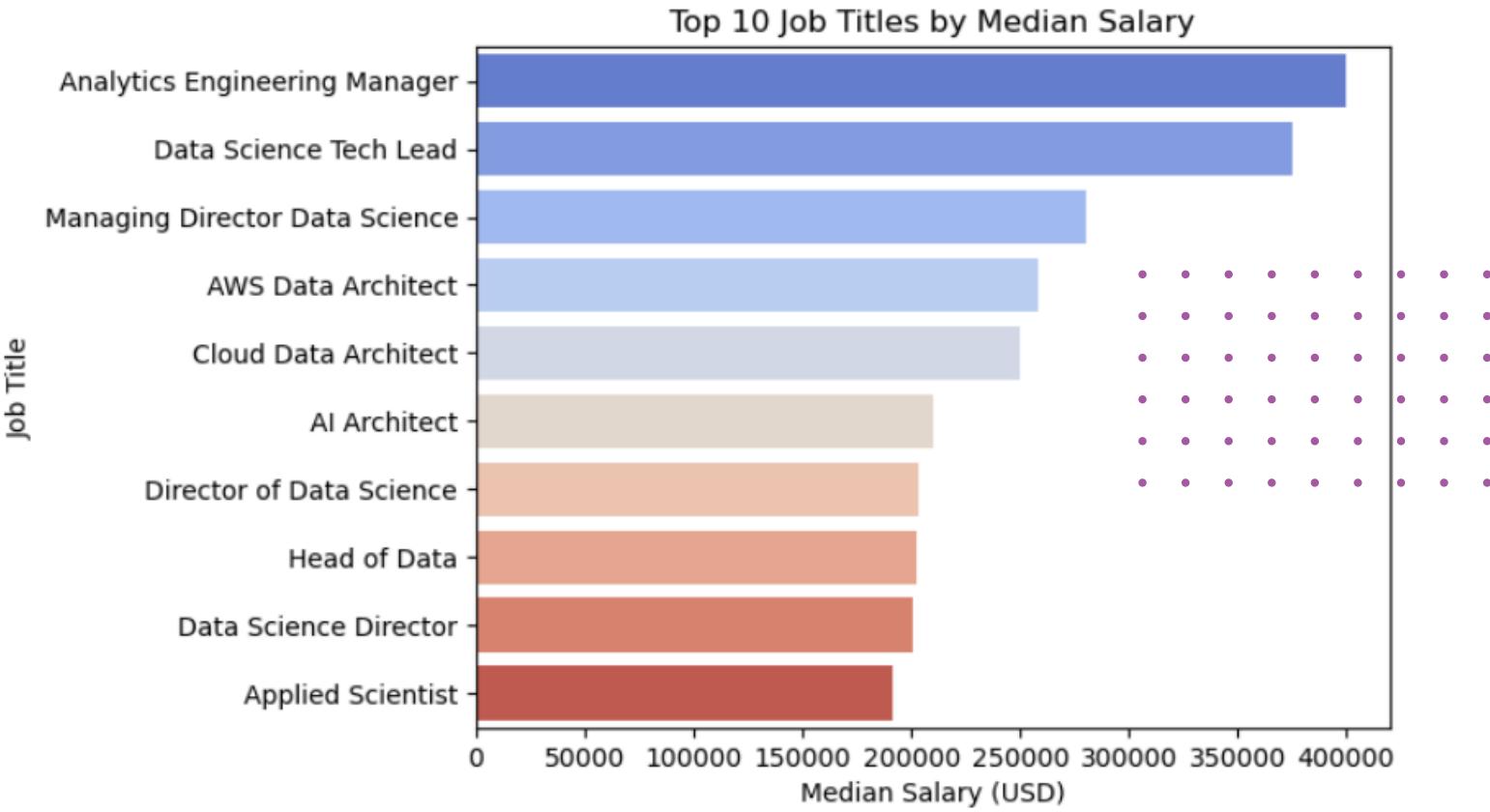
SALARY DISTRIBUTION



The salary distribution is right-skewed, meaning most people earn low to medium salaries, while a few earn very high salaries.

As a Business Analyst, this means using the median salary is more accurate and meaningful than the average, because the average can be pulled up by a few very high salaries.

SALARIES BY JOB TITLES



Roles like Analytics Engineering Manager and Data Science Tech Lead offer the highest median salaries.

Business Analyst Recommendation

This suggests that if a company wants to hire top-level data talent, especially for advanced roles like ML Engineers or Data Architects, they need to allocate a higher budget for these positions.

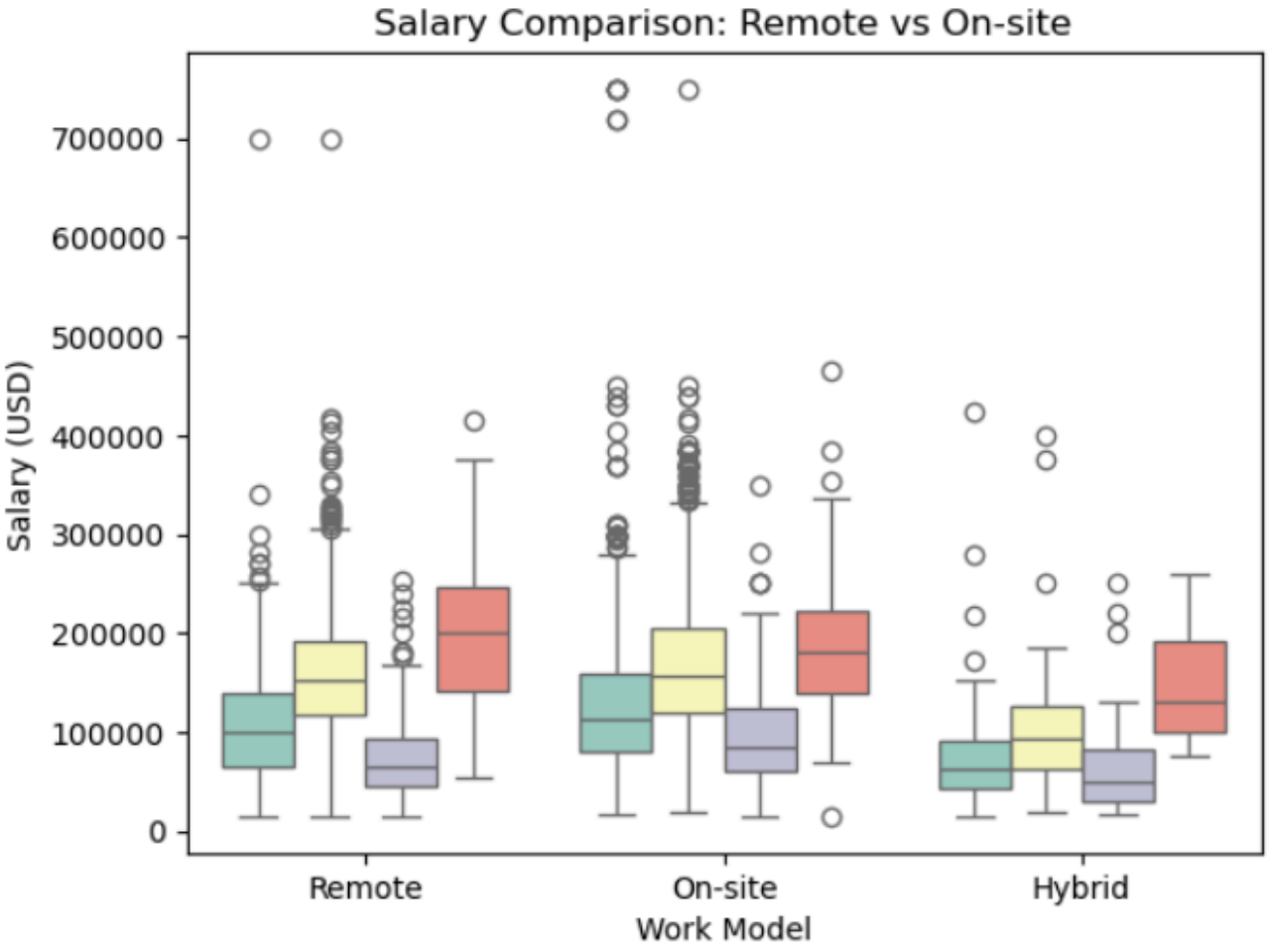
SALARIES BY EXPERIENCE LEVEL

Salaries increase sharply with experience. Executive-level professionals earn much more than entry-level ones.

Business Analyst Recommendation

1. Plan future budgets smartly. Companies should forecast payroll expenses as employees gain more experience.
2. Retain senior talent. Since senior professionals are costly to replace, offer them benefits like leadership roles, bonuses, or equity.
3. Upskill mid-level staff. Invest in training programs to promote from within and avoid expensive external hiring.

REMOTE VS ON-SITE SALARY



Remote workers seem to earn slightly more on average than on-site or hybrid workers. This could mean companies are hiring high-paid international talent remotely.

Business Analyst Recommendation

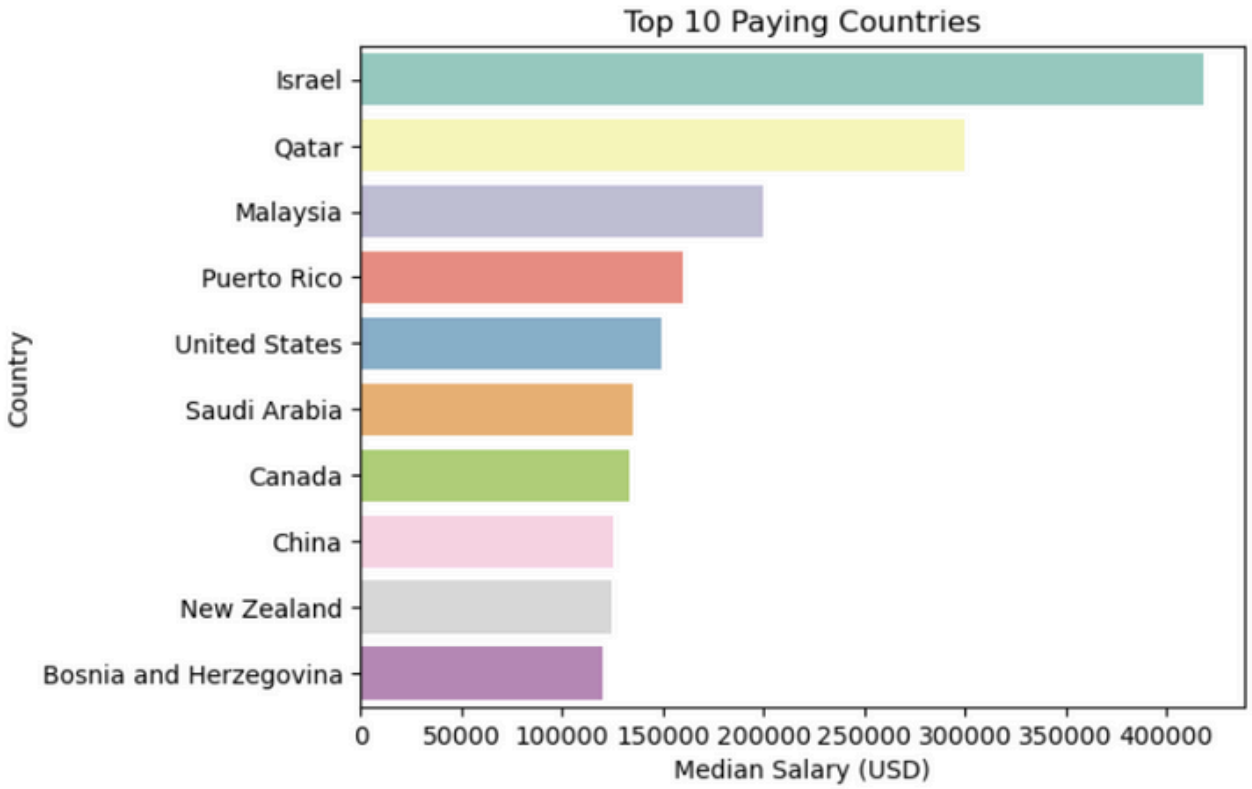
1. Analyze how remote hiring affects overall salary budgets, especially in global hiring.
2. Evaluate if higher remote salaries bring better ROI in terms of skills and productivity.
3. Suggest a balanced hiring policy that combines local and global talent while managing costs.

TOP 10 PAYING COUNTRIES

Countries like Israel, Qatar, and Malaysia offer the highest median salaries in the data science field. If companies are hiring talent from these countries, they need to carefully evaluate the return on investment.

Business Analyst Recommendation

Hiring from high-paying countries should only be considered when the candidate's skills and value match the higher salary. Businesses should assess if the employee will contribute through innovation, performance, or leadership to balance the cost. This helps in managing budgets while still attracting top talent.

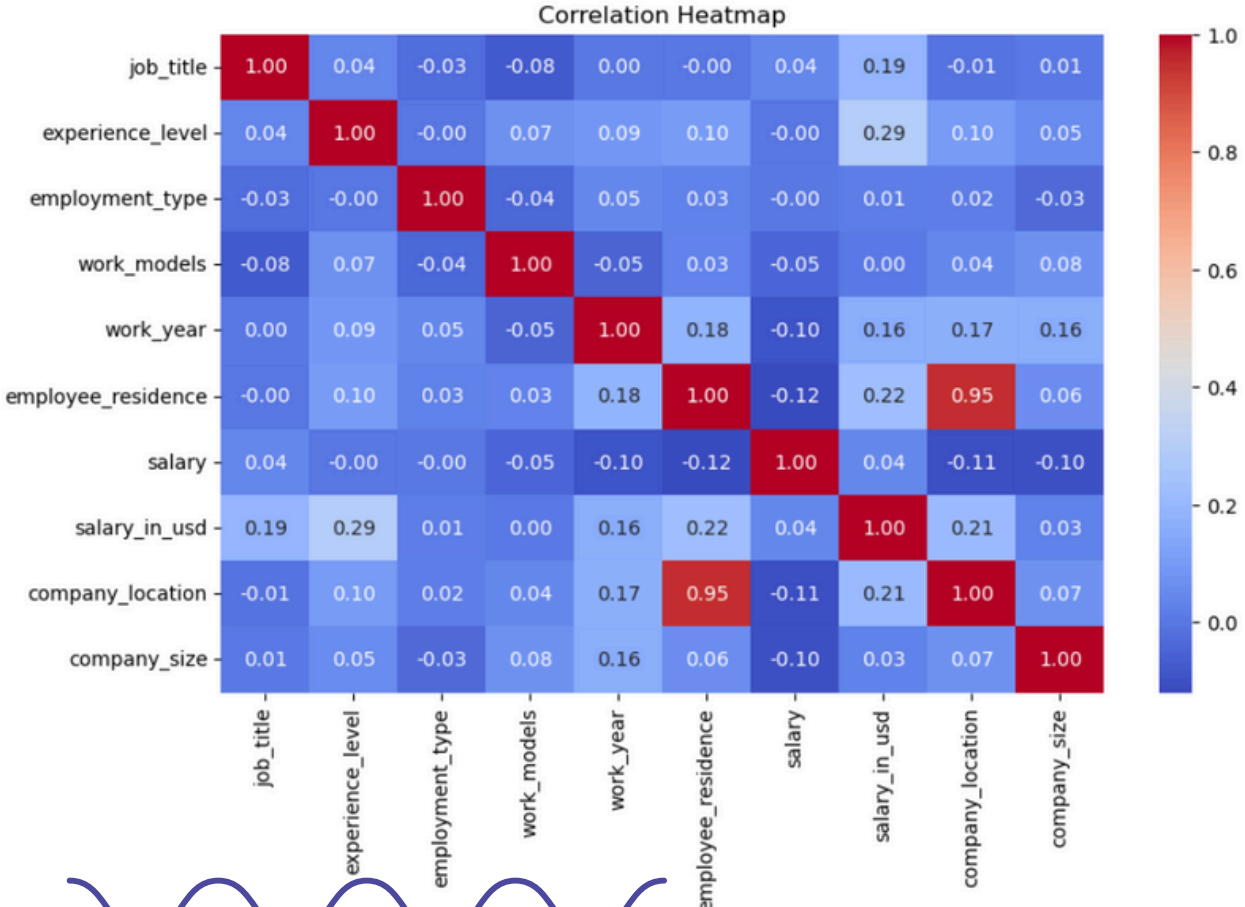


SALARY PLANNING INSIGHTS

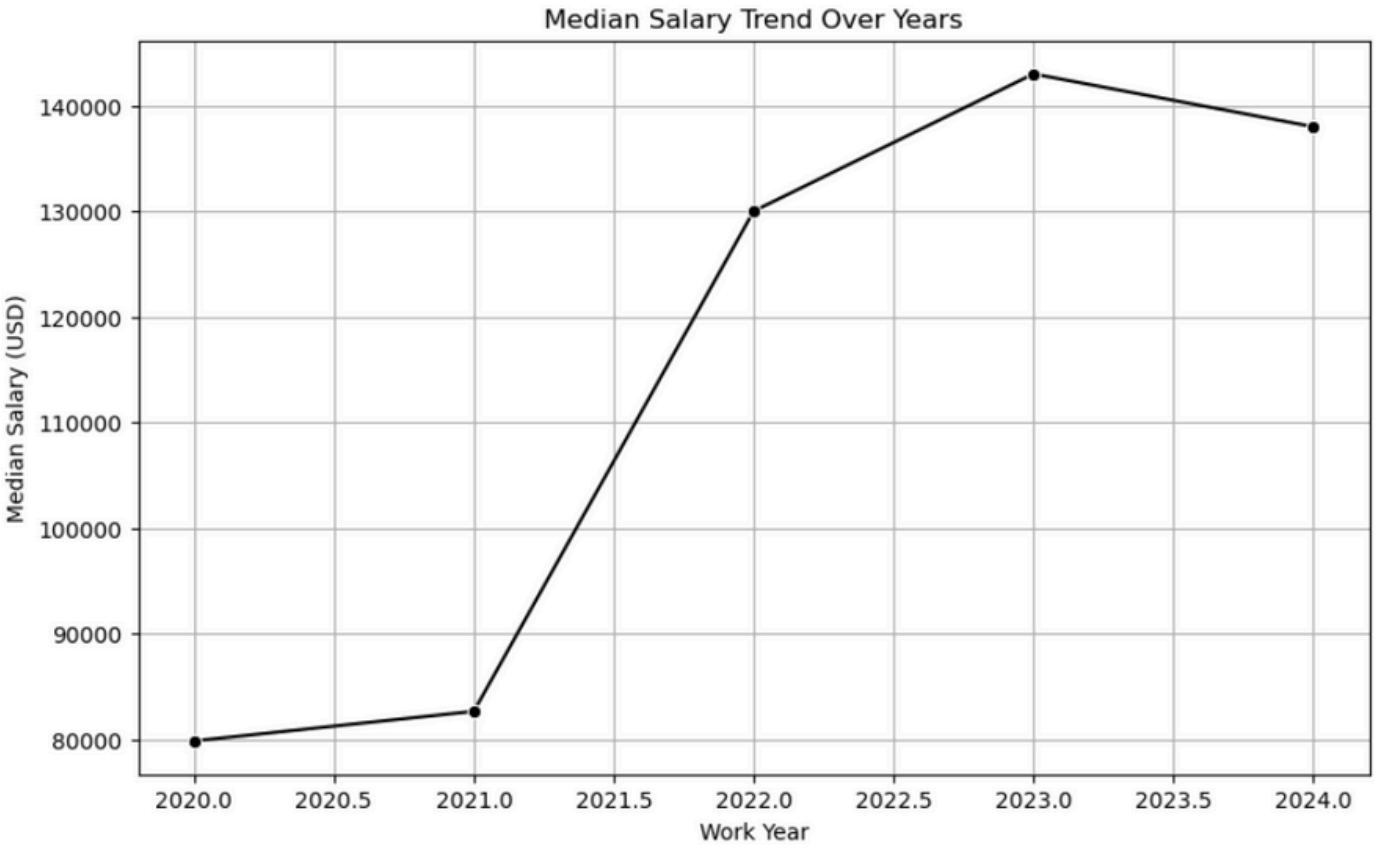
The heatmap shows that experience level and job title have a moderate positive relationship with salary in USD. This means as experience increases or job roles become more advanced, salary also increases.

Business Analyst Recommendation

While planning salaries, companies should focus more on experience level and job title as these factors affect salary the most. Location and year also have some influence, but employment type and work model show very weak impact. So, for budgeting and pay structure, experience and role should be the main focus.



SALARY TRENDS OVER YEAR



The line graph shows that median salaries in the data science field increased strongly from 2020 to 2023, rising from $80,000 to $143,000. The biggest jump occurred between 2021 and 2022, showing a rapid rise in demand for data professionals during that time. However, in 2024, there is a slight decline to $138,000, suggesting a possible market correction or saturation in the job market.

Business Analyst Recommendation

Companies should consider aligning salary structures with market trends. The sharp growth between 2021 and 2023 signals a need for competitive offers during tech growth phases. However, the dip in 2024 highlights the importance of skill-based pay and smart budgeting. Businesses are advised to invest more in upskilling current employees, focus on retention strategies, and regularly benchmark salaries against industry standards to avoid overpaying or losing top talent.

RECOMMENDATIONS AS BUSSINESS ANALYST

For Data Professionals

- Target Strategic Roles: Focus on roles like Tech Lead, Data Architect, and Analytics Manager that offer >$300K salaries.
- Choose Medium-Sized Firms: They offer the best pay-to-growth ratio, with fewer hierarchies and faster decision cycles.
- Invest in Skills That Pay: Master cloud platforms (AWS, GCP), ML Engineering, and data pipelines,directly linked to top salaries.
- Go Remote with Strategy: Remote roles offer high upside and target companies in Israel, US, or Qatar for global salary benchmarks.
- Build a Hybrid Brand: Combine technical skills with business storytelling to transition from data analyst to decision-maker.

For Employers & HR Leaders

- Benchmark Like Medium Firms: Re-evaluate compensation, mid-sized companies are outcompeting large firms in talent acquisition.
- Promote Tech Leadership: Salaries peak with data leadership roles, invest in upskilling mid-level staff to Tech Leads and Architects.
- Optimize Remote Hiring: Remote workers deliver high ROI and create structured onboarding, clear KPIs, and performance-linked pay.
- Close the Pay Gap: Address gender and location bias by standardizing skill-based pay bands and regular comp audits.
- Retain with Career Velocity: Offer fast-track growth and leadership mentoring to prevent attrition at senior levels.

For Educators & Policymakers

- Modernize Curriculum: Align with industry and focus on production ML, cloud tools, and real business use cases.
- Train for Global Work: Equip talent for freelancing and remote-first roles with client-facing skills and cross-cultural fluency.
- Upskill Mid-Career Talent: Launch executive programs to transition experienced professionals into high-paying data leadership roles.

CODE FOR EDA

PHASE 1

```
# Importing required libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy import stats
from sklearn.preprocessing import MinMaxScaler
```

```
df['salary_in_usd'] = pd.to_numeric(df['salary_in_usd'], errors='coerce')
df['salary_in_usd'].describe()
```

```
count      6599.000000
mean     145560.558569
std       70946.838070
min       15000.000000
25%       95000.000000
50%      138666.000000
75%      185000.000000
max      750000.000000
Name: salary_in_usd, dtype: float64
```

```
df['salary_currency'].unique()
```

```
array(['USD', 'GBP', 'EUR', 'TRY', 'THB', 'ZAR', 'SGD', 'INR', 'PLN',
       'PHP', 'NOK', 'ILS', 'HKD', 'AUD', 'CHF', 'CAD', 'BRL', 'JPY',
       'HUF', 'DKK', 'CLP', 'MXN'], dtype=object)
```

PHASE 2

```
# median salary by company size
median_salary = df.groupby('company_size')['salary_in_usd'].median().reset_index()


sns.barplot(
    data=median_salary,
    x='company_size',
    y='salary_in_usd',
    hue='company_size',
    palette='pastel',
    legend=False
)
plt.title("Median Salary by Company Size")
plt.xlabel("Company Size")
plt.ylabel("Median Salary (USD)")
plt.show()
```

CODE FOR EDA

PHASE 3

```
#Salary Distribution
# Histogram (bars)

sns.histplot(df['salary_in_usd'], bins=30, kde=True, color="purple" )

plt.title("Salary Distribution")
plt.xlabel("Salary (USD)")
plt.ylabel("Frequency")
plt.show()
```

PHASE 4

```
#Median Salary by Job Title
#bar plot
median_by_job = df.groupby("job_title")["salary_in_usd"].median().sort_values(ascending=False).head(10)

df_jobs = median_by_job.reset_index()
df_jobs.columns = ['job_title', 'median_salary']

sns.barplot(data=df_jobs, y='job_title', x='median_salary', hue='job_title', palette='coolwarm', legend=False)
plt.title("Top 10 Job Titles by Median Salary")
plt.xlabel("Median Salary (USD)")
plt.ylabel("Job Title")
plt.show
```

PHASE 5

```
#Salary by Experience Level
#bar plot
median_by_exp = df.groupby("experience_level")["salary_in_usd"].median().sort_values()

df_exp = median_by_exp.reset_index()
df_exp.columns = ['experience_level', 'median_salary']

sns.barplot(data=df_exp, x='experience_level', y='median_salary', hue='experience_level', palette='Set3', legend=False)
plt.title("Median Salary by Experience Level")
plt.xlabel("Experience Level")
plt.ylabel("Median Salary (USD)")
plt.show()
```

CODE FOR EDA

PHASE 6

```
#Remote vs On-site Salary Comparison
#box plot

sns.boxplot(x="work_models", y="salary_in_usd", data=df, hue='experience_level', palette='Set3', legend=False)
plt.title("Salary Comparison: Remote vs On-site")
plt.xlabel("Work Model")
plt.ylabel("Salary (USD)")
plt.show()
```

PHASE 7

```
top_countries = df.groupby("employee_residence")["salary_in_usd"].median().sort_values(ascending=False).head(10)

# Convert to DataFrame
top_df = top_countries.reset_index()
top_df.columns = ['employee_residence', 'median_salary']

# bar plot
sns.barplot(data=top_df, x='median_salary', y='employee_residence', hue='employee_residence', palette='Set3', legend=False)
plt.title("Top 10 Paying Countries")
plt.xlabel("Median Salary (USD)")
plt.ylabel("Country")
plt.show()
```

PHASE 8

```
#Salary by Experience Level
#bar plot
median_by_exp = df.groupby("experience_level")["salary_in_usd"].median().sort_values()

df_exp = median_by_exp.reset_index()
df_exp.columns = ['experience_level', 'median_salary']

sns.barplot(data=df_exp, x='experience_level', y='median_salary', hue='experience_level', palette='Set3', legend=False)
plt.title("Median Salary by Experience Level")
plt.xlabel("Experience Level")
plt.ylabel("Median Salary (USD)")
plt.show()
```

CODE FOR EDA

PHASE 9

```
print(df.dtypes)
```

```
job_title             object
experience_level      object
employment_type       object
work_models           object
work_year              int64
employee_residence    object
salary                 int64
salary_currency       object
salary_in_usd          int64
company_location      object
company_size          object
dtype: object
```

```
# Convert categorical columns to numerical using label encoding
df_encoded = df.copy()
categorical_cols = ['job_title', 'experience_level', 'employment_type', 'work_models',
                    'employee_residence', 'company_location', 'company_size']

for col in categorical_cols:
    df_encoded[col] = df_encoded[col].astype('category').cat.codes


correlation_matrix = df_encoded.corr(numeric_only=True)

#heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
```

PHASE 10

```
#salary trends over years
#line plot
plt.figure(figsize=(10, 6))
sns.lineplot(x='work_year', y='salary_in_usd', data=salary_trend, marker='o', color='black')
plt.title("Median Salary Trend Over Years")
plt.xlabel("Work Year")
plt.ylabel("Median Salary (USD)")
plt.grid(True)
plt.show()
```