Employee Well-Being and Work Flexibility in the Technology Industry (2025)

- Explore how work arrangements, salary, and workload influence employee mental, physical, and social well-being.
- Focus on technology industry roles, analyzing patterns in burnout, social isolation, health issues, and work-life balance.
- Goal: Identify key factors affecting employee health and guide data-driven workplace strategies.

Presented by: Lee Sze Ning

Data Cleaning & Data Transformation:

- Handle missing values in both datasets (e.g., missing prices, order dates, or last visit dates).

```
[4]: # check for missing values for employee
employee.isna().sum()
```

```
[4]: Survey_ID           0
Survey_Date         0
Age                 0
Gender              0
Region              5
Job_Role            0
Work_Arrangement    0
Salary_Range        0
dtype: int64
```

```
[5]: # removal of categorical missing values (since it only involved less than 1% in total
employee = employee.dropna(subset=['Region'])

# Double check
print(employee['Region'].isna().sum())
```

```
0
```

```
[6]: # check for missing values for health
health.isna().sum()
```

```
[6]: Survey_ID                  0
Hours_Per_Week             14
Mental_Health_Status      157
Burnout_Level               0
Work_Life_Balance_Score     0
Physical_Health_Issues     49
```

Data Joining:

- Join the datasets using a unique identifier.
- Perform groupby to uncover relationships between variables.

```
5]: merged_df = pd.merge(employee, health, on='Survey_ID',how='inner')
merged_df['Mental_Health_Issues'] = merged_df['Mental_Health_Status'].apply(
    lambda x: 'Yes' if x != 'No Issue Reported' else 'No'
)
merged_df = merged_df.rename(columns={'Physical_Health_Issues': 'Physical_Health_Status'})
merged_df['Physical_Health_Issues'] = merged_df['Physical_Health_Status'].apply(
    lambda x: 'Yes' if x != 'No Issue Reported' else 'No'
)
merged_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 588 entries, 0 to 587
Data columns (total 18 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Survey_ID               588 non-null    object
 1   Survey_Date             588 non-null    datetime64[ns]
 2   Age                     588 non-null    int64
 3   Gender                  588 non-null    object
 4   Region                  588 non-null    object
 5   Job_Role                588 non-null    object
 6   Work_Arrangement        588 non-null    object
 7   Salary_Range            588 non-null    object
 8   Salary_Midpoint($)      588 non-null    int64
 9   Hours_Per_Week          588 non-null    float64
 10  Mental_Health_Status    588 non-null    object
 11  Burnout_Level           588 non-null    object
 12  Work_Life_Balance_Score 588 non-null    int64
 13  Physical_Health_Status  588 non-null    object
 14  Social_Isolation_Score  588 non-null    int64
 15  Workload_Level          588 non-null    object
```

Data Manipulation and Data Transformation:

- Ensure data types and formatting are consistent.
- Create new columns that are helpful for data analysis

```
9]: summary_employee = []

for col in employee.columns:
    unique_vals = employee[col].unique()
    summary_employee.append({
        'Column': col,
        'Num_Unique': employee[col].nunique(),
        'Sample_Unique_Values': unique_vals[:5]  # first 5 unique va
    })

summary_employee_df = pd.DataFrame(summary_employee)

# Display the summary table
print(summary_employee_df)
```

```
             Column  Num_Unique  \
0         Survey_ID         588
1       Survey_Date          26
2               Age          44
3            Gender           4
4            Region           6
5          Job_Role          24
6  Work_Arrangement           3
7      Salary_Range           5

                          Sample_Unique_Values
0      [SURV009, SURV012, SURV013, SURV038, SURV040]
1  [2025-06-01, 2025-06-02, 2025-06-03, 2025-06-0...
2                          [30, 64, 25, 42, 29]
3      [Female, Male, Non-binary, Prefer not to say]
4  [South America, North America, Oceania, Europe...
```

```
#Thirs Big Question

# 1. Filter for High Work-Life Balance (score = 5)
high_wlb = merged_df[merged_df['Work_Life_Balance_Score'] == 5]

# 2. Count High WLB occurrences by Job Role and take top 10
top_roles = high_wlb['Job_Role'].value_counts().head(10)

# 3. Highlight role
colors = ['tomato' if role == 'Sales Representative' else 'skyblue' for role in top_roles.index]

# 4. Plot pie chart
plt.figure(figsize=(8, 8))
wedges, texts, autotexts = plt.pie(
    top_roles,
    labels=top_roles.index,
    autopct='%1.1f%%',
    colors=colors,
    startangle=140,
    textprops={'fontsize': 10}
)

plt.title('Top 10 Job Roles: High Work-Life Balance Distribution')

plt.tight_layout()
plt.show()
```

4️⃣ Exploratory Data Analysis

Descriptive Analysis:

- Explore overall descriptive analysis.
- Filter subsets to answer big questions.

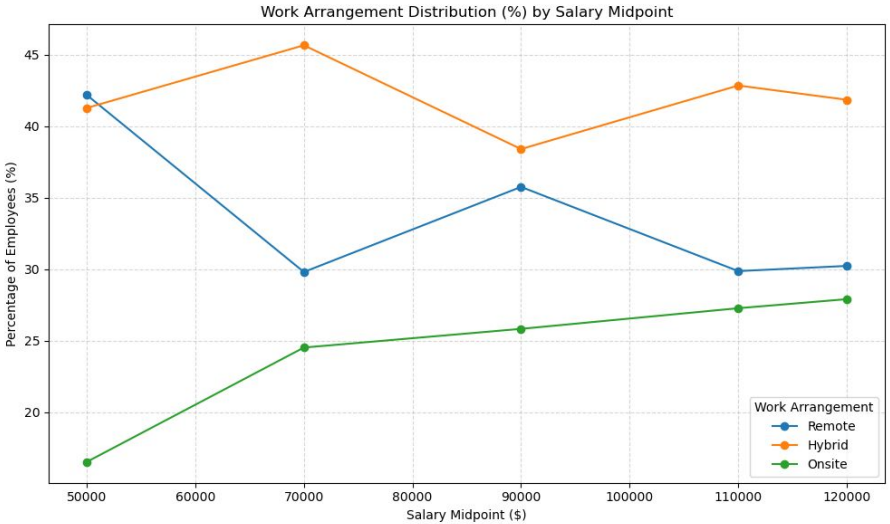
```
]: #describe all the numberical variables
merged_df.describe()
```

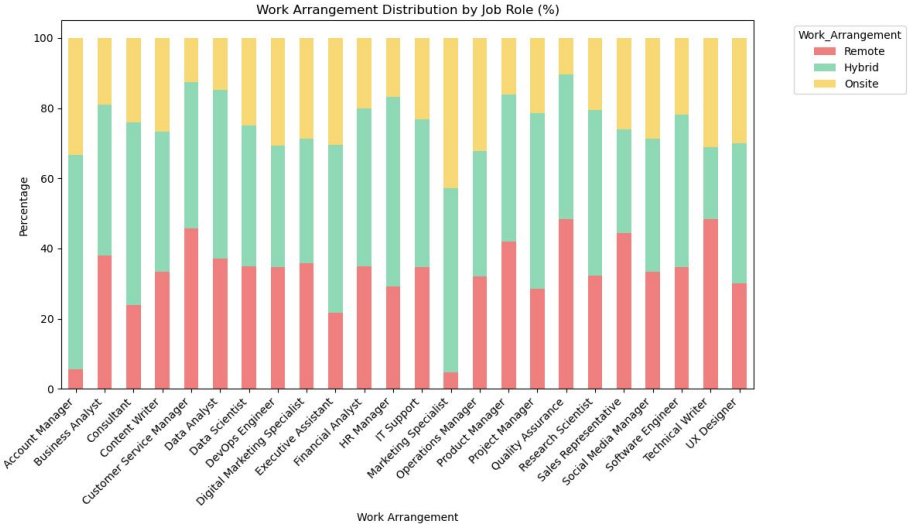
	Survey_Date	Age	Salary_Midpoint($)	Hours_Per_Week	Work_Life_Balance_Score	Social_Isolation_Score
count	588	588.000000	588.000000	588.000000	588.000000	588.000000
mean	2025-06-13 11:13:28.163265280	43.517007	80323.129252	49.383978	2.955782	2.775510
min	2025-06-01 00:00:00	22.000000	50000.000000	35.000000	1.000000	1.000000
25%	2025-06-07 00:00:00	32.000000	70000.000000	42.000000	2.000000	2.000000
50%	2025-06-14 00:00:00	44.000000	70000.000000	49.412781	3.000000	3.000000
75%	2025-06-20 00:00:00	55.000000	90000.000000	56.000000	4.000000	4.000000
max	2025-06-26 00:00:00	65.000000	120000.000000	65.000000	5.000000	5.000000
std	NaN	12.820928	21527.629940	8.509108	1.160722	1.169143

```
']: #describe all the categorical variables
merged_df.describe(include='object')
```

Big Question 1 – Salary & Work Arrangment

Question: How do Remote, Hybrid, and Onsite arrangements relate to salary and job role?





Key Findings:

- Hybrid is most common across salary levels; **Remote dominates below $50k.**
- By job role: **Customer Service, QA, Technical Writers** have the highest Remote share.

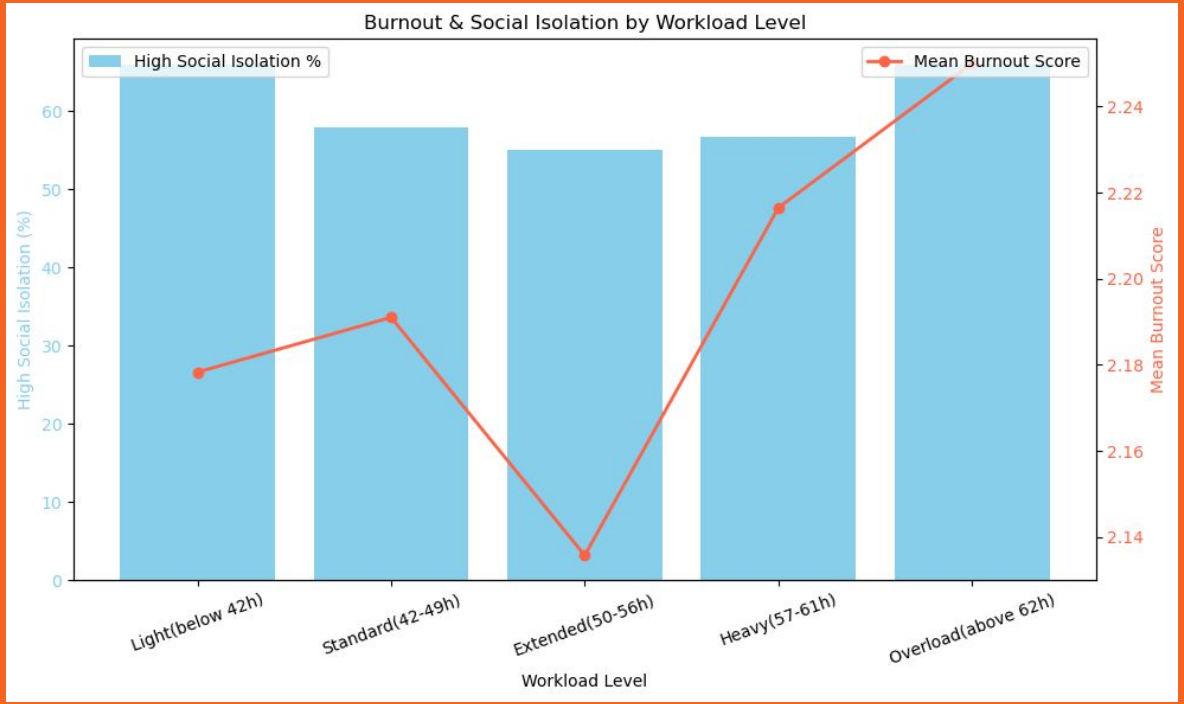
Flexibility varies by salary and role, with some lower-paid roles offering more remote options!

Big Question 2 – Working hours & Well-Being

Question: How do working hours affect well-being?

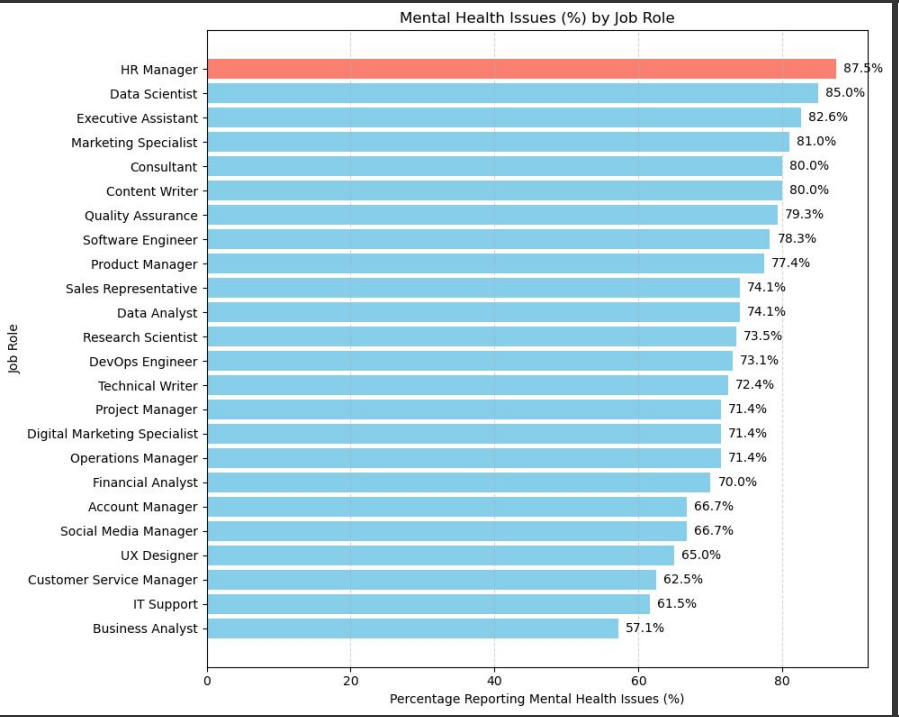
- ➔ **Burnout** rises with workload, slightly lower at 50–56h/week.
- ➔ **Social Isolation** lowest at 50–56h/week.

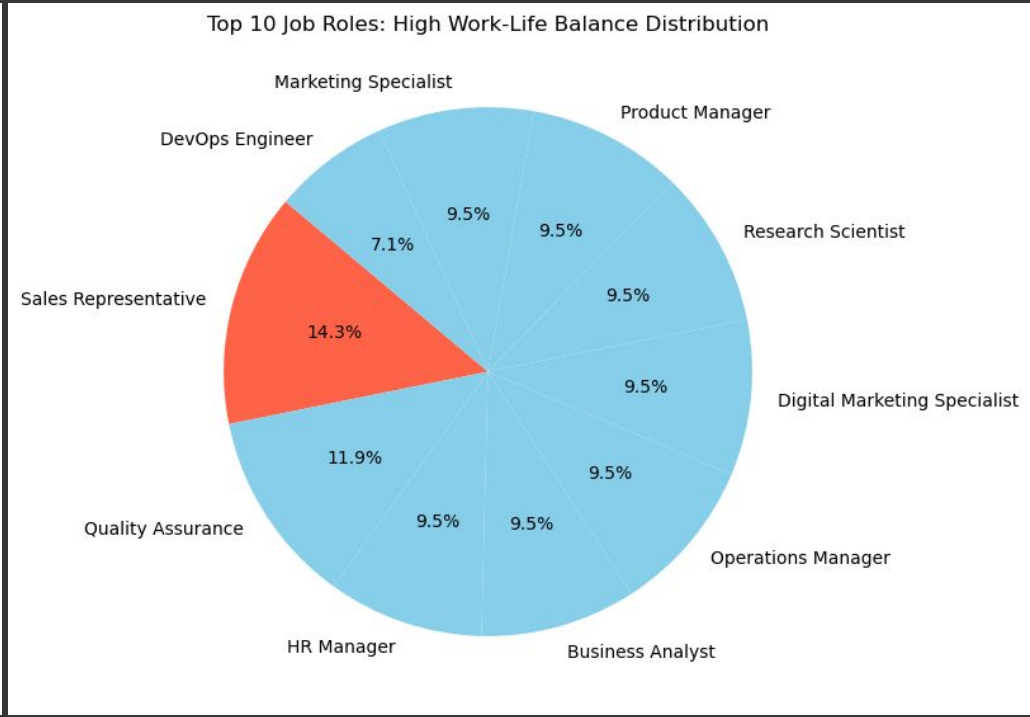
High workload and long hours negatively impact health, but mid-range hours may improve social engagement.



Presented by: Lee Sze Ning

— Big Question 3 : Job Role Impact Question: How do job roles influence well-being?





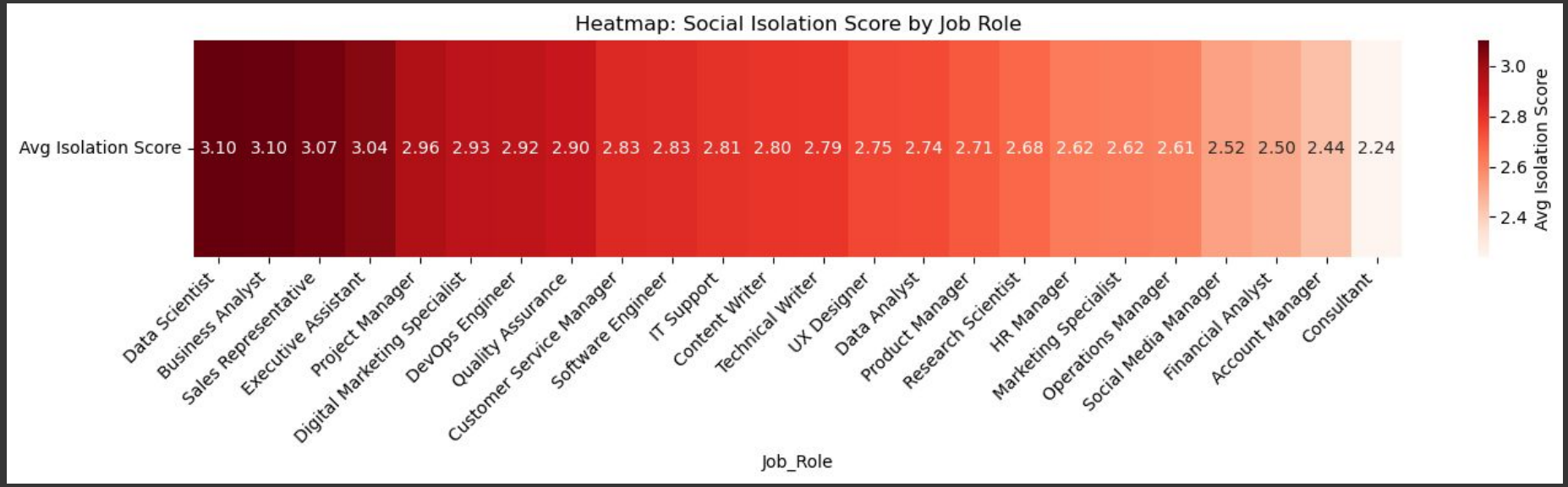
Mental Health:

HR Managers (87.5%) & Data Scientists (85%) most affected.

Work-Life Balance:

Sales Reps lead in high scores.

Big Question 3 : Job Role Impact



Social Isolation: Highest in Data Scientists, lowest in Consultants.

Insight

Employee well-being is strongly **role-dependent** across mental, physical, and social dimensions.

Presented by: Lee Sze Ning

Summary & Implications

Takeaways:

- **Salary & role** shape access to flexible work arrangements.
- **Working hours** strongly influence burnout, social isolation, and health.
- **Job roles** are key predictors of employee well-being.

Implications:

- Optimize **workload distribution** and flexible arrangements.
- Target **high-risk roles** for health support initiatives.
- Use insights to guide **policy and workplace strategy**.

Presented by: Lee Sze Ning