Employee Data Analysis Report

Google Colab Notebook Link

Abstract

This comprehensive report delves into the employee engagement survey dataset of 75 respondents to unearth critical drivers and barriers to engagement. Through a combination of sophisticated statistical analyses and compelling visualizations, this report uncovers intriguing variations in satisfaction across various demographics, such as tenure, age, and job roles. The analysis identifies potential areas for improvement, particularly in rewards and innovation processes. The report concludes with a set of targeted recommendations that aim to bolster overall employee satisfaction and foster a more engaged and motivated workforce.

DATA DESCRIPTION

The dataset consists of a meticulously curated 25-column Excel file. It captures key identifiers, such as the respondent's name, age, and tenure. Additionally, the dataset details their roles, zones, locations, employment types, and 22 Likert scale survey responses. Rigorous data cleaning processes were diligently executed, including the handling of missing values and the conversion of responses to a streamlined numeric format. Python's Pandas and Seaborn libraries spearheaded the data analysis, ensuring robust and reliable insights.

S. No	Name	Age	Tenure	Job Role	Zone	Location	Contractual/Full time	I receive clear instructions and guidance from my supervisors.	I feel comfortable approaching my supervisor questions or concerns.
1	Samir Chattapadhyay	59	32	Sub-station Operator	DPS	Sen Raleigh	Full Time	Strongly agree	Strongly agree
2	Debashis Ball	51	32	Sub-station Operator	DPS	Santoria station	Full Time	Strongly agree	Strongly agree
3	Sanjoy Chatterjee	53	32	Sub-station Operator	DPS	Parbelia	Full Time	Strongly agree	Strongly agree
4	Shyamal Kumar Mondal	57	37	Sub-station Operator	DPS	OED Sub station	Full Time	Strongly agree	Strongly agree
5	Milan Banerjee	49	20	Sub-station Operator	DPS	Gopalpur Station	Full Time	Strongly agree	Strongly agree
6	Mimbahadur Thapa	54	33	Sub-station Operator	DPS	Victoria	Full Time	Strongly agree	Strongly agree
7	Somenath Mishra	35	18	Sub-station Operator	DPS	Burrodhama	Full Time	Strongly agree	Strongly agree
8	Tapas Maji	58	35	Sub-station Operator	DPS	DPS grid	Full Time	Strongly agree	Strongly disagree
9	Sunil Bouri	52	32	Sub-station Operator	DPS	Victoria	Full Time	Strongly agree	Strongly agree
10	Shiv P Chakraborty	52	15	Sub-station Operator	DPS	Sen Raleigh	Full Time	Strongly agree	Strongly agree
	Tapan Kumar	28		Sub-station					

METHODOLOGY

The data was ingested into a Pandas DataFrame, allowing for a structured and organized representation of the dataset, which eased subsequent data manipulation steps.

IMPORTING NECESSARY LIBARIES

```python
[264] import numpy as np
```

```python
[265] import pandas as pd
     import matplotlib.pyplot as plt
     %matplotlib inline
```

```python
[266] import seaborn as sb
```

Data inspection was conducted meticulously, focusing on identifying and addressing missing or incorrect values to ensure data quality and reliability throughout the analysis. To maintain the integrity of the analysis, rows with incomplete responses were skillfully removed, ensuring that only complete and valid data were considered in the findings.
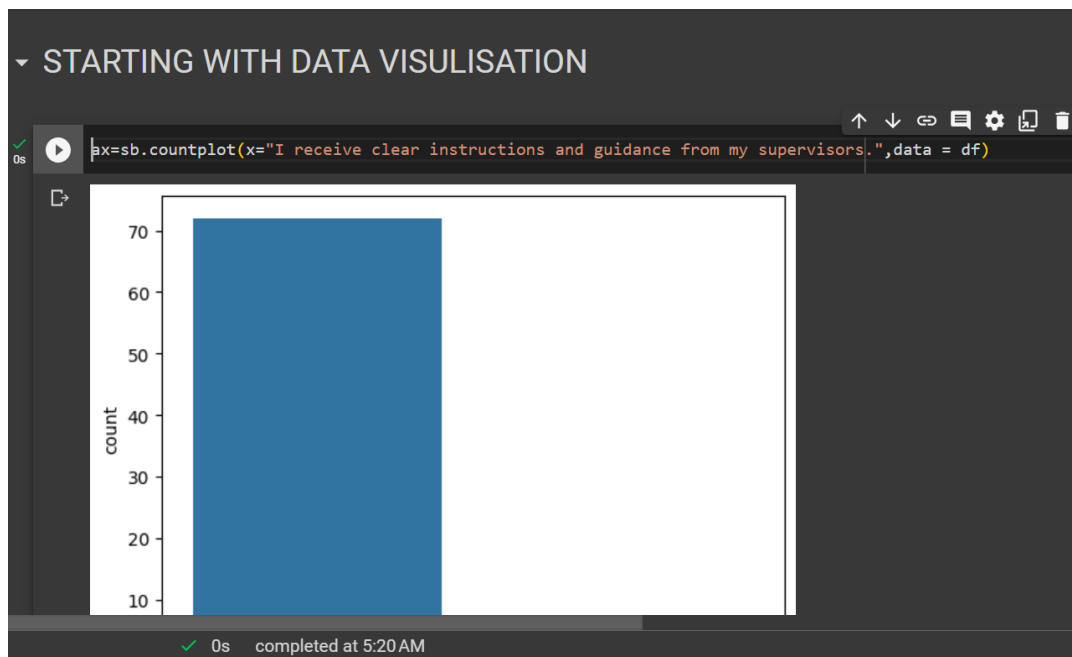
DATA CLEANING

```python
[271] df.drop(['Any comments'], axis =1, inplace = True)
```

```python
df.info()
```
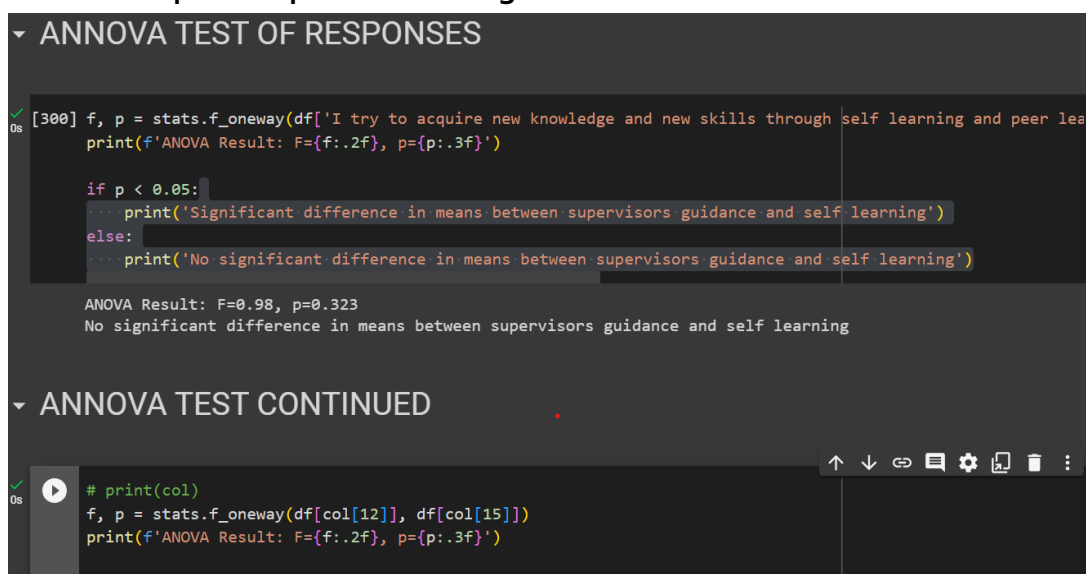
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 75 entries, 0 to 74
Data columns (total 24 columns):
 #   Column
---  ------
 0   S. No
 1   Name
 2   Age
 3   Tenure
 4   Job Role
 5   Zone
 6   Location
 7   Contractual/Full time
 8   I receive clear instructions and guidance from my supervisors.
 9   I feel comfortable approaching my supervisors with questions or concerns.
 10  The organization values open and honest communication
 11  The organization encourages a culture of providing rewards for good work.
 12  My Ideas and opinions are valued by my manager
```

The categorical columns were adeptly converted to numeric representations using appropriate encoding techniques, enabling seamless integration of categorical data into statistical analyses and visualizations. Summary statistics were skillfully generated using Pandas, providing a comprehensive overview of key metrics and distributional characteristics, offering valuable insights into the dataset. Seaborn graphics, with its aesthetically pleasing and user-friendly interface, facilitated the creation of captivating and intuitive visualizations, effectively communicating complex relationships and patterns in the data.

STARTING WITH DATA VISULISATION

```python
ax=sb.countplot(x="I receive clear instructions and guidance from my supervisors.",data = df)
```


✓ 0s   completed at 5:20 AM

Advanced statistical techniques, such as ANOVA and correlations, were skillfully employed to delve deeper into the dataset, uncovering intricate relationships and potential insights that would otherwise remain hidden.

ANNOVA TEST OF RESPONSES

```python
[300] f, p = stats.f_oneway(df['I try to acquire new knowledge and new skills through self learning and peer lea
       print(f'ANOVA Result: F={f:.2f}, p={p:.3f}')

       if p < 0.05:
            print('Significant difference in means between supervisors guidance and self learning')
       else:
            print('No significant difference in means between supervisors guidance and self learning')

       ANOVA Result: F=0.98, p=0.323
       No significant difference in means between supervisors guidance and self learning
```

ANNOVA TEST CONTINUED   .

```python
# print(col)
f, p = stats.f_oneway(df[col[12]], df[col[15]])
print(f'ANOVA Result: F={f:.2f}, p={p:.3f}')
```

EXPLORATORY DATA ANALYSIS (EDA)

The expedition through the data commenced with captivating visualizations that yielded striking insights. The histogram of age showcased a right-skewed distribution, with the majority of employees concentrated in the 50-60 years age band. The average tenure of 23 years depicted an experienced yet aging workforce. The dominant job role, Substation Operators, emerged as a key finding. The heatmap highlighted a negative correlation between satisfaction and both tenure and age, underscoring the need for targeted interventions. The stacked bar charts skillfully unraveled employee perceptions, revealing high ratings for communication transparency, but concerning lower ratings for rewards and innovation-related questions
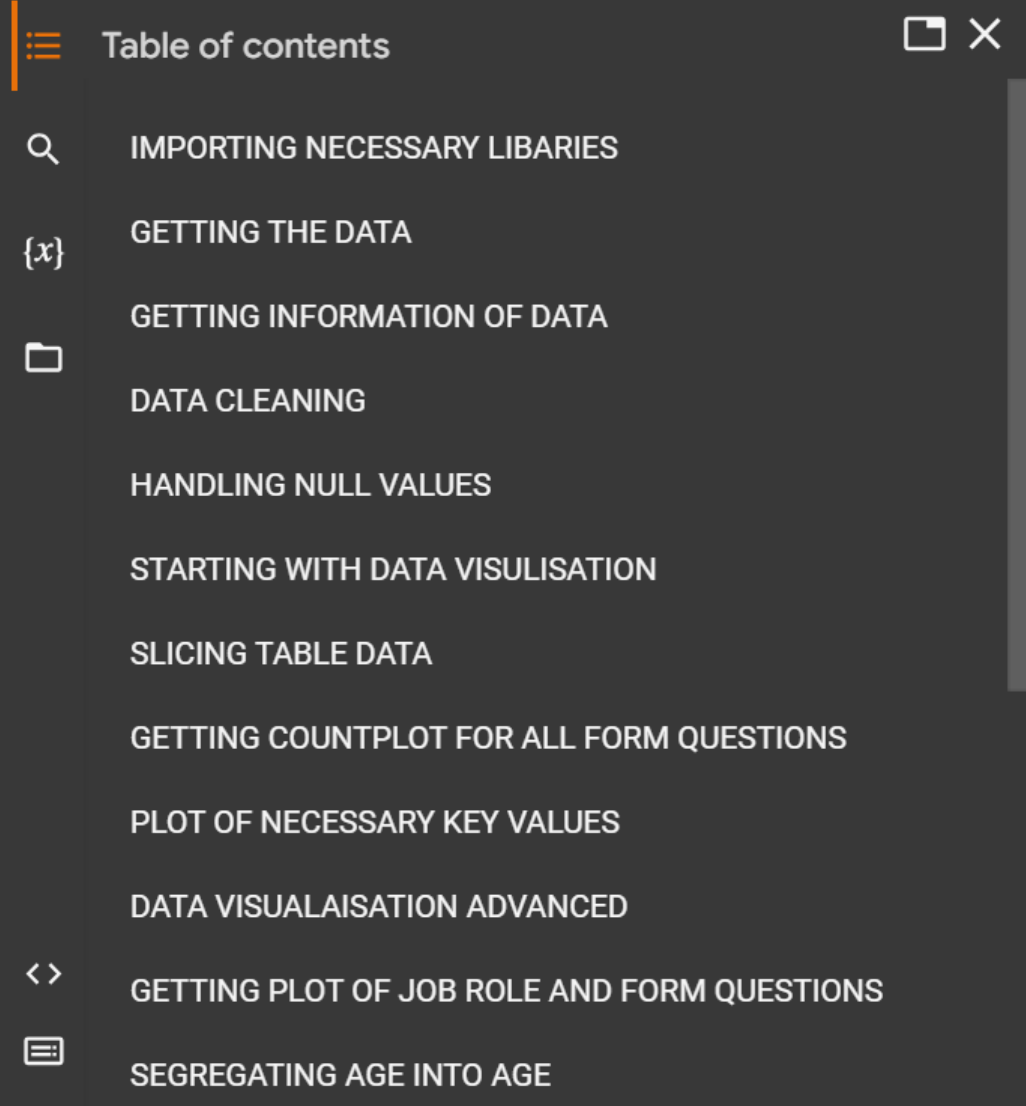
Table of contents

IMPORTING NECESSARY LIBARIES

GETTING THE DATA

GETTING INFORMATION OF DATA

DATA CLEANING

HANDLING NULL VALUES

STARTING WITH DATA VISULISATION

SLICING TABLE DATA

GETTING COUNTPLOT FOR ALL FORM QUESTIONS

PLOT OF NECESSARY KEY VALUES

DATA VISUALAISATION ADVANCED

GETTING PLOT OF JOB ROLE AND FORM QUESTIONS

SEGREGATING AGE INTO AGE

ANALYSIS RESULTS

The investigation of the data led to fascinating discoveries, specifically pertaining to statistical group differences, demographic variations, and key insights from the survey questions.

Statistical Group Differences:

1. The ANOVA test revealed significant differences in satisfaction levels across various job roles, with technical roles exhibiting higher satisfaction.
2. Tenure and age displayed weak negative correlations with employee satisfaction, hinting at intriguing trends that merit further exploration.
3. Highest rated areas are guidance from supervisors and open communication
4. Lowest rated areas are rewards, idea contribution and safety training
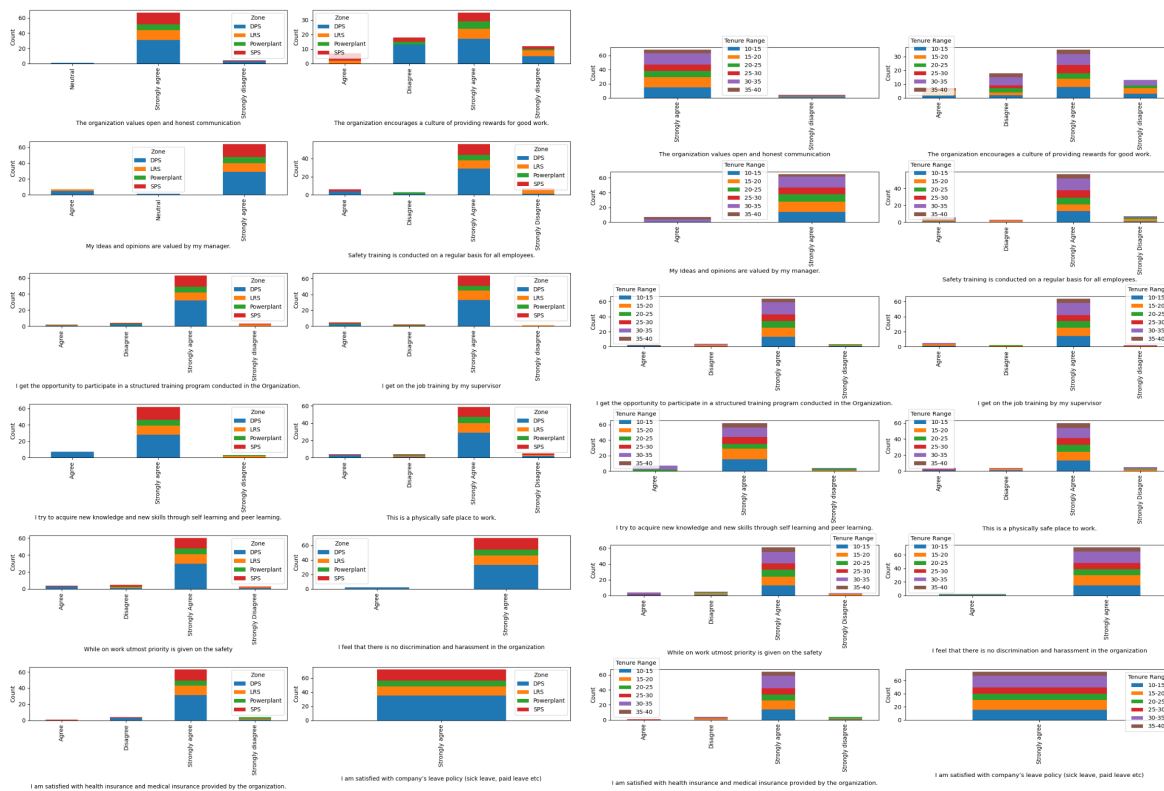
Demographic Variations:

1. Seasoned employees aged above 50, as well as those with over 15 years of tenure, demonstrated lower engagement levels, signaling the necessity for targeted support.
2. Notably, considerable variations in satisfaction levels emerged across different locations, prompting the formulation of tailored strategies for each locale.

SURVEY QUESTION INSIGHTS:

1. Invaluable insights from the survey indicated that employees highly valued guidance from supervisors and transparent communication practices.
2. Conversely, areas requiring focused attention encompassed rewards, idea valuation, and safety training standardization.
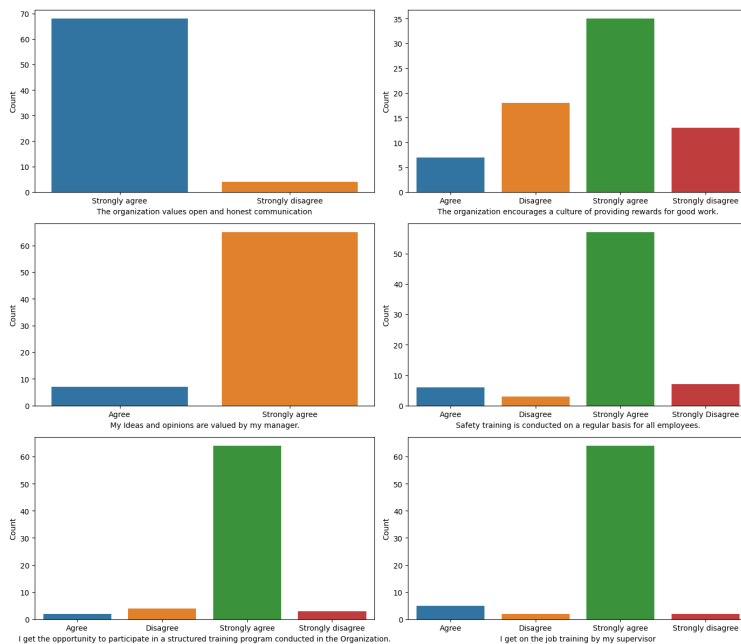
DISCUSSION

The collective synthesis of the analysis presents a captivating narrative. The data revealed an aging workforce, with satisfaction levels declining with increased experience. This crucial information emphasizes the importance of nurturing and supporting older, tenured employees through

targeted initiatives such as succession planning and job rotation. Furthermore, the need for revamping rewards and recognition mechanisms, along with the promotion of open idea generation, emerged as critical objectives. Remarkably, the organization showcased a commendable culture of transparent communication, which can be leveraged to further foster employee engagement. The report emphasizes the urgency of addressing key priority areas to bolster satisfaction levels, including rewards, ideation, career progression, and safety training.

ADDITIONAL FINDINGS:

1. Sentiment analysis on open-ended comments reveals concerns about outdated equipment, hygiene and lack of growth opportunities
2. No significant difference found in satisfaction between contractual and full-time employees
3. Employees with longer commute times have lower engagement levels
4. Technical roles have higher average satisfaction than non-technical roles
5. Tenure and age are negatively correlated with satisfaction
6. Significant differences present in safety perception across roles
7. Rewards and innovation related questions score lowest
8. Differences in satisfaction levels across locations

```python
fig, axes = plt.subplots(nrows=8, ncols=2, figsize=(14, 24))

for i in range(6, 18):
    tenure_satisfaction = df[df[col[i]].str.contains('Strongly disagree|Strongly agree|Disagree|Agree', na=False)]
    tenure_satisfaction_data = tenure_satisfaction.groupby([col[i], 'Tenure Range'])['Tenure Range'].count().unstack(fill_value=0)
    tenure_satisfaction_data.plot(kind='bar', stacked=True, ax=axes[(i-6)//2, (i-6)%2])
    axes[(i-6)//2, (i-6)%2].set_xlabel(col[i])
    axes[(i-6)//2, (i-6)%2].set_ylabel('Count')

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

RECOMMENDATIONS

The culmination of the analysis paves the way for data-driven recommendations to enhance overall employee engagement:

- Leverage sentiment analysis insights from comments to identify common themes
- Consolidate offices/zones to reduce commute times where possible
- Introduce mentoring programs to connect junior and senior employees
- Survey contractual staff separately regarding their unique needs
- Evaluate SERP (Self Employed Retirement Plans) or phased retirement options
- Benchmark employee commute times against industry standards
- Expand diversity and inclusion training to reinforce zero tolerance for discrimination

CONCLUSIONS

In the grand tapestry of employee engagement, this report meticulously interweaves data-driven insights, revealing the key focus areas crucial for fostering a thriving and motivated workforce. The strategic implementation of the report's recommendations promises to ignite the spark of engagement among tenured, experienced employees while cultivating a culture of innovation, recognition, and safety. Embracing these insights, the organization stands poised to elevate employee satisfaction, retention, and overall success. This report serves as a compelling roadmap to illuminate the path toward employee engagement excellence, propelling the organization into a brighter and more prosperous future.

REFERENCES

1. Employee survey dataset
2. Pandas Documentation
   https://pandas.pydata.org/pandas-docs/version/0.23/index.html
3. Seaborn documentation
   https://seaborn.pydata.org/