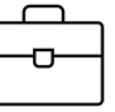


STAGE 1 FINAL PROJECT

KELOMPOK: Pengabdi FinPro



Group Members



- 1 Faldo Aditya
- 2 Putri
- 3 Virgo Gilang Pratama (Leader)
- 4 Rizki Anugrah Pratama
- 5 Rizky Nur Ichsan
- 6 Raihan Hadi Syahputra
- 7 Rizky Putra Okfradifa
- 8 Nira Aulia Hanifah



#BACKGROUND

Attrition karyawan merupakan karyawan yang berniat meninggalkan perusahaan secara sukarela. Hal ini didorong oleh banyak faktor yang utama adalah ketidak terpenuhinya ekspektasi akan kompensasi kerja berupa gaji atau bonus [1] dan ketidakpuasan dengan perusahaannya baik berupa pekerjaan, lingkungan maupun hubungan sosial dengan atasan dan rekan kerjanya [2][3].

Kami sebagai tim HC Data Scientist akan membuat model yang dapat memprediksi attrition karyawan secara keseluruhan maupun individu. Selengkapnya dapat melihat repository tim kami [4].

Resources:

- [1] [A Study On Causes Of Employee Attrition](#)
- [2] [An Empirical Study of Employee Attrition In Select Professional Institute of Lucknow](#)
- [3] [The effect of work environment, stress, and job satisfaction on employee turnover intention](#)
- [4] [Repository Pengabdi FinPro.](#)



#DESCRIPTIVE ANALYTICS

```
Age          0  
Attrition    0  
BusinessTravel 0  
DailyRate     0  
Department    0  
DistanceFromHome 0  
Education      0  
EducationField 0  
EmployeeCount   0  
EmployeeNumber   0  
EnvironmentSatisfaction 0  
Gender         0  
HourlyRate     0  
JobInvolvement 0  
JobLevel       0  
JobRole        0  
JobSatisfaction 0  
MaritalStatus   0  
MonthlyIncome    0  
MonthlyRate     0  
NumCompaniesWorked 0  
Over18         0  
OverTime        0  
PercentSalaryHike 0  
PerformanceRating 0  
RelationshipSatisfaction 0  
StandardHours    0  
StockOptionLevel 0  
TotalWorkingYears 0  
TrainingTimesLastYear 0  
WorkLifeBalance 0  
YearsAtCompany   0  
YearsInCurrentRole 0  
YearsSinceLastPromotion 0  
YearsWithCurrManager 0  
dtype: int64
```

Finding: Tidak ada value yang missing dan terduplikasi pada dataset ini.

```
df.duplicated()  
  
0      False  
1      False  
2      False  
3      False  
4      False  
...  
1465    False  
1466    False  
1467    False  
1468    False  
1469    False  
Length: 1470, dtype: bool  
  
duplicates = df[df.duplicated()]  
print(duplicates)  
  
Empty DataFrame  
Columns: [Age, Attrition, BusinessTravel, DailyRate, Department, DistanceFromHome, Education, EducationFi  
eld, EmployeeCount, EmployeeNumber, EnvironmentSatisfaction, Gender, HourlyRate, JobInvolvement, JobLevel  
, JobRole, JobSatisfaction, MaritalStatus, MonthlyIncome, MonthlyRate, NumCompaniesWorked, Over18, OverTi  
me, PercentSalaryHike, PerformanceRating, RelationshipSatisfaction, StandardHours, StockOptionLevel, Total  
WorkingYears, TrainingTimesLastYear, WorkLifeBalance, YearsAtCompany, YearsInCurrentRole, YearsSinceLast  
Promotion, YearsWithCurrManager]  
Index: []  
  
[0 rows x 35 columns]  
  
No Duplicated value has been found
```



#DESCRIPTIVE ANALYTICS

```
df['EmployeeCount'].describe()
```

```
count    1470.0
mean      1.0
std       0.0
min      1.0
25%      1.0
50%      1.0
75%      1.0
max      1.0
Name: EmployeeCount, dtype: float64
```

```
df['StandardHours'].describe()
```

```
count    1470.0
mean     80.0
std      0.0
min     80.0
25%     80.0
50%     80.0
75%     80.0
max     80.0
Name: StandardHours, dtype: float64
```

```
df['EmployeeNumber'].describe()
```

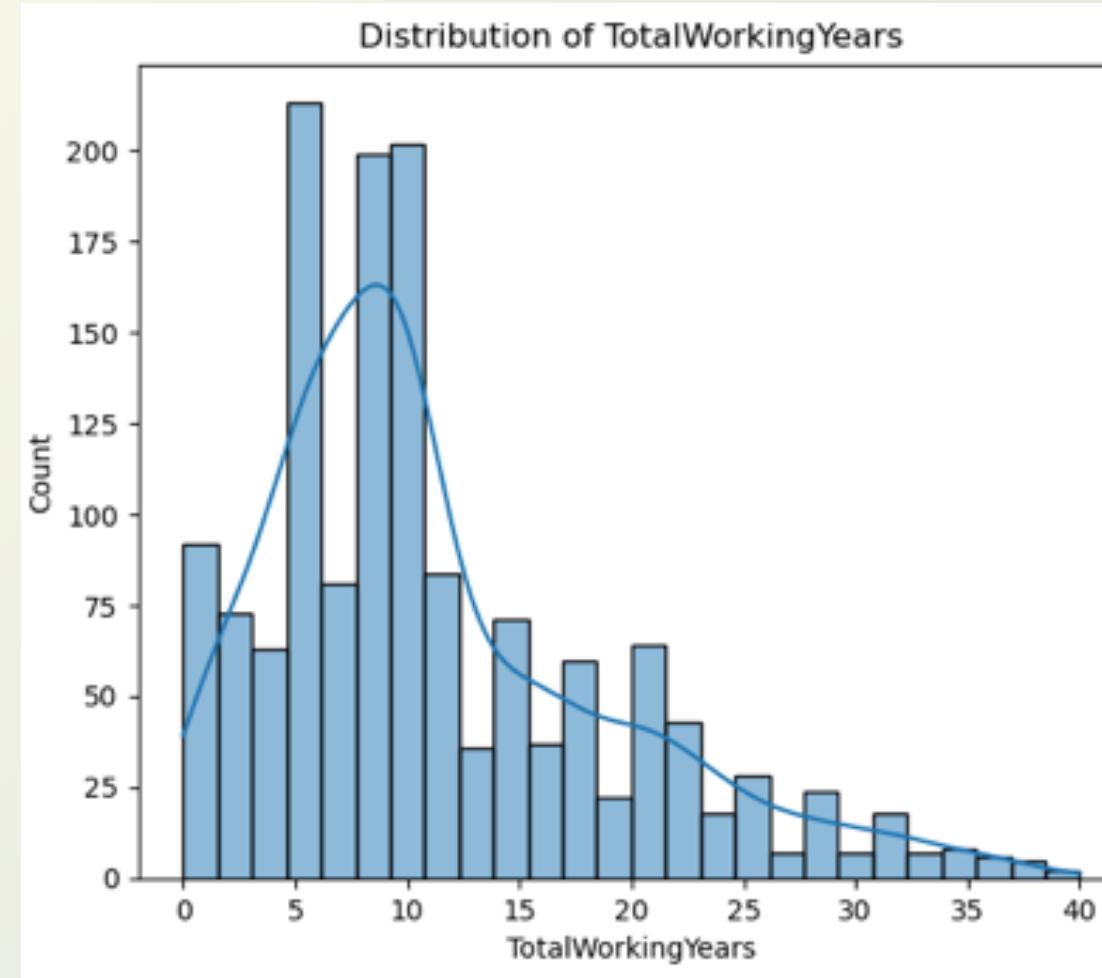
```
count    1470.000000
mean    1024.865306
std     602.024335
min     1.000000
25%    491.250000
50%    1020.500000
75%    1555.750000
max    2068.000000
Name: EmployeeNumber, dtype: float64
```

Finding: Data describe tidak menunjukkan nilai yang aneh kecuali 'EmployeeCount', 'StandardHours' dan 'EmployeeNumber'. Fitur 'EmployeeCount' dan 'StandardHours' hanya memiliki 1 unique value saja. Sementara, untuk 'EmployeeNumber' kami berasumsi value ini merujuk pada ID employee. Di mana angka di dataset yang lebih besar daripada total karyawan yang ada saat ini bermakna bahwa employee ID yang tidak terdata adalah mereka yang sudah resign.

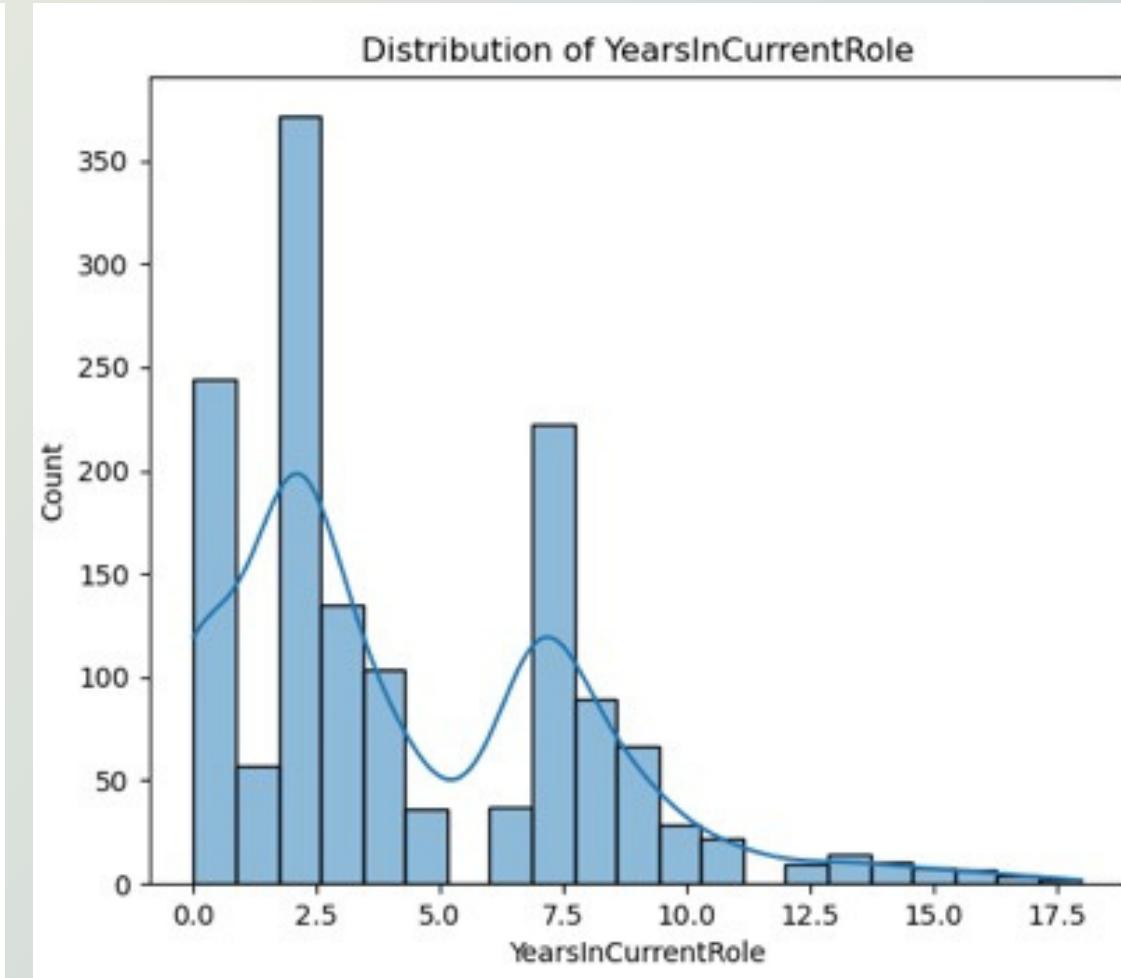
#UNIVARIATE ANALYSIS



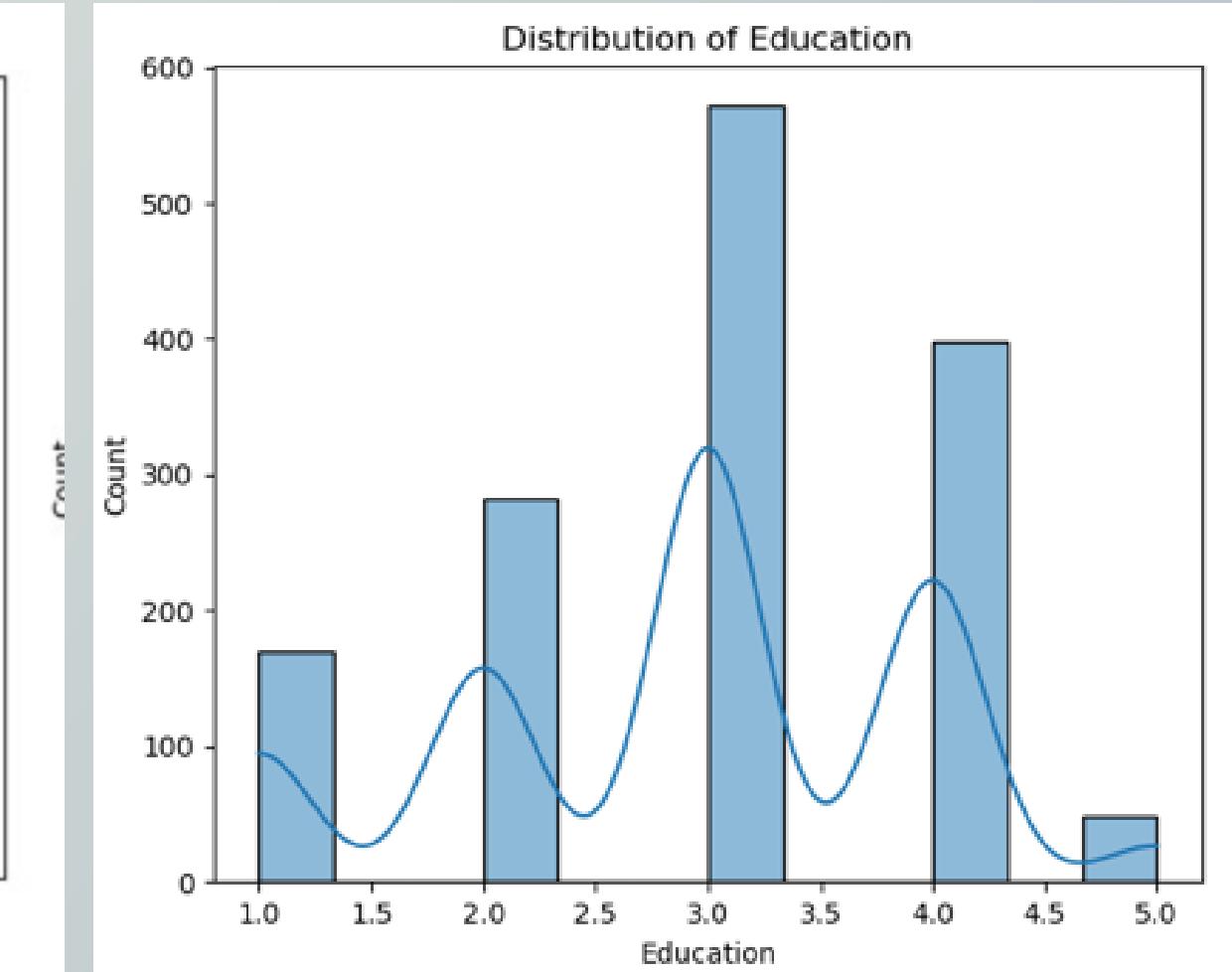
Unimodal



Bimodal



Multimodal

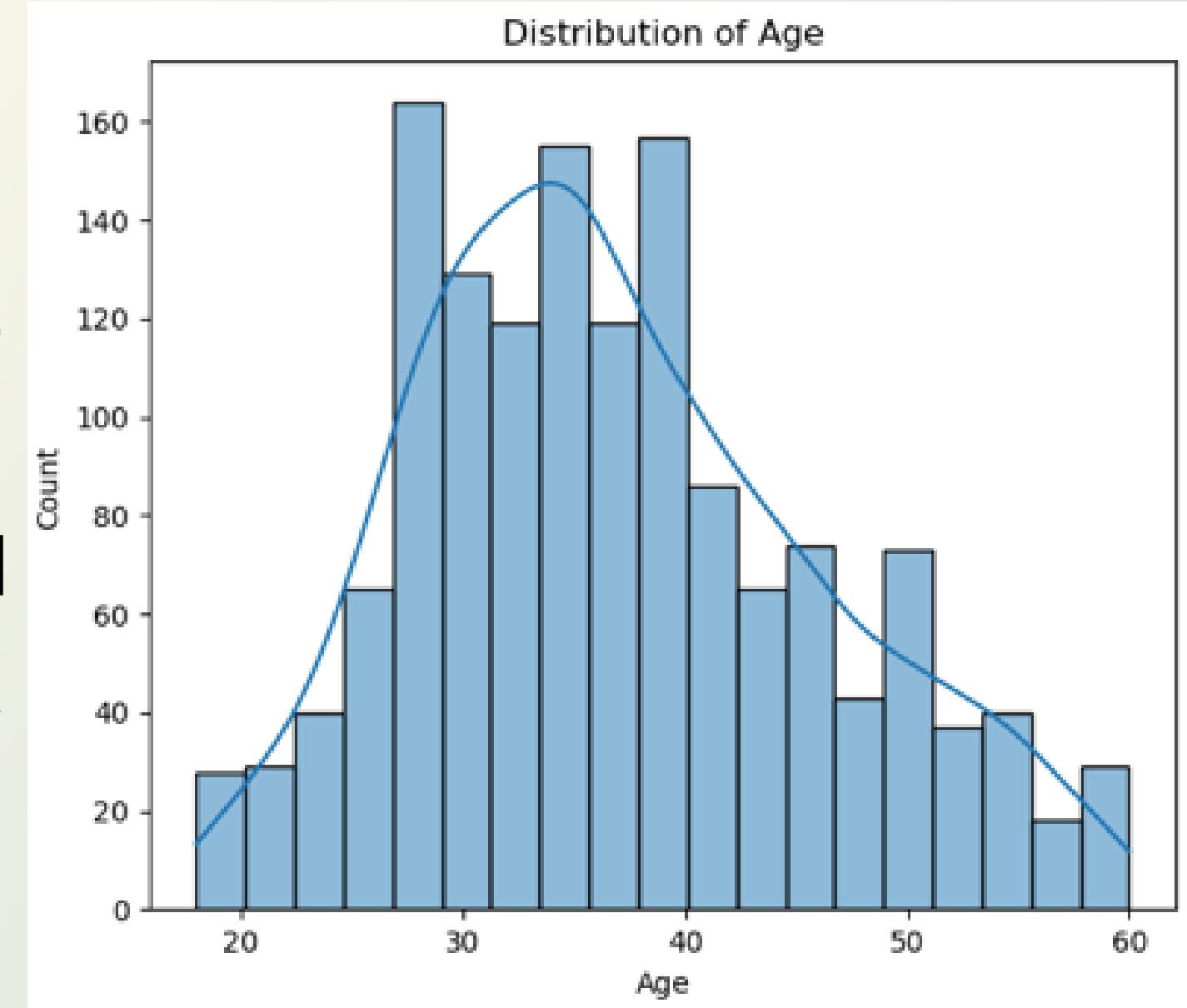


Dalam dataset ini terdapat 3 jenis distribusi berdasarkan jumlah peaknya. Unimodal (1 peak) seperti pada fitur TotalWorkingYears, Bimodal (2 peak) seperti fitur YearsInCurrentRole dan Multimodal (>2 peak) seperti fitur Education



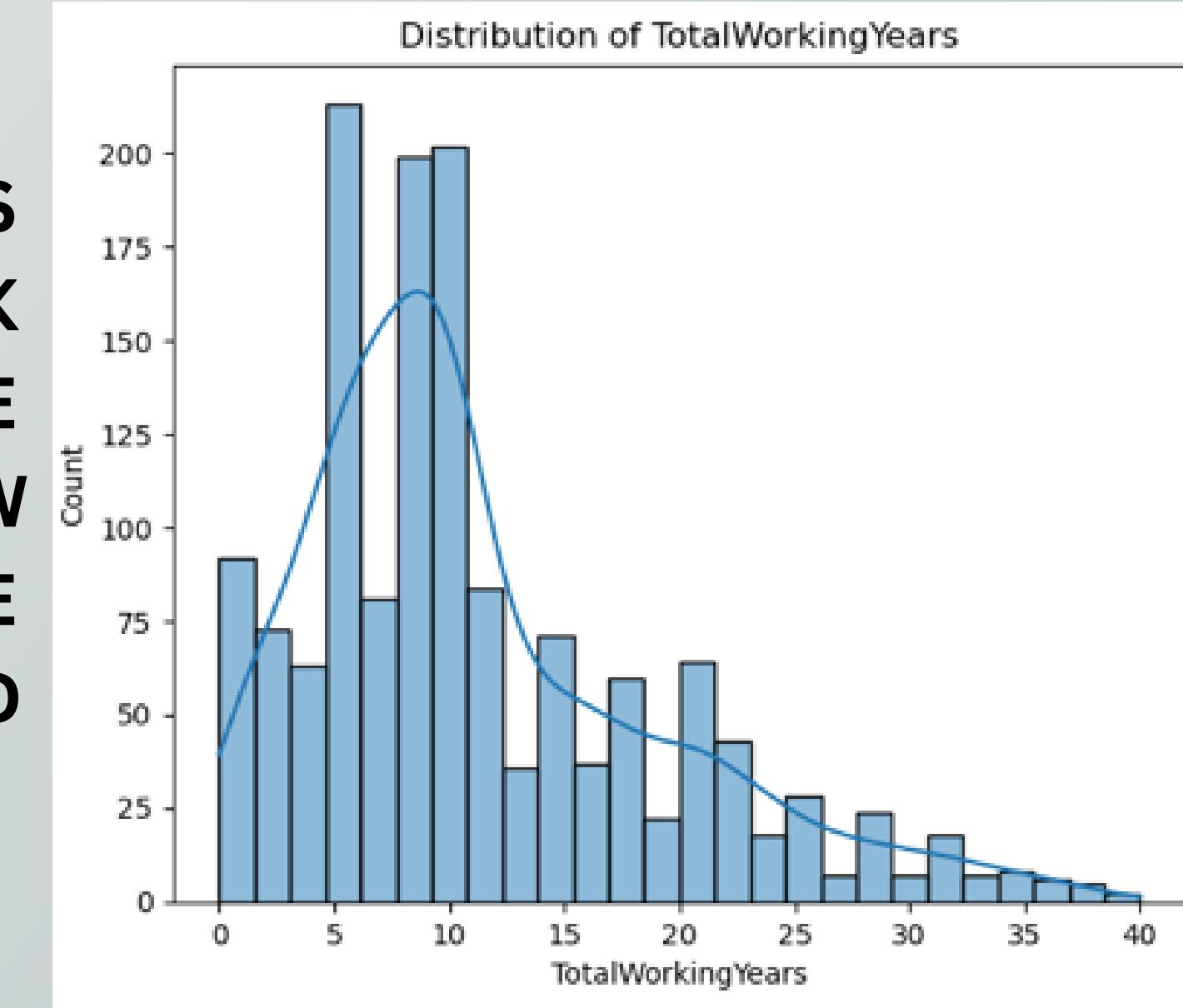
#UNIVARIATE ANALYSIS

N
O
R
M
A
L



Terdapat pula tipe distribusi normal seperti pada fitur Age.

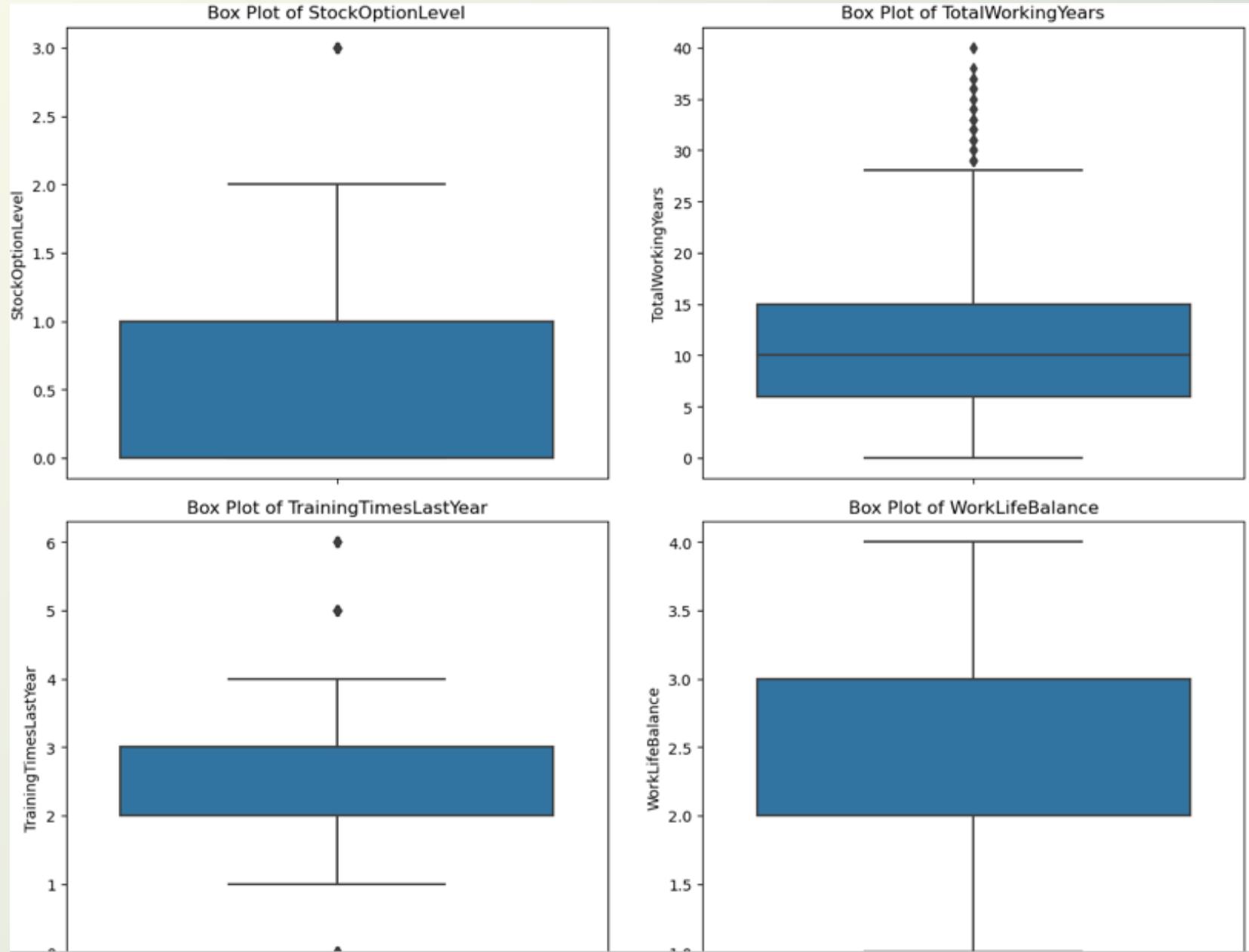
P
O
S
I
T
I
V
E
L
Y



Terdapat beberapa fitur dengan tipe distribusi positively skewed atau berekor ke kanan. Salah satu contohnya adalah fitur TotalWorkingYears



#UNIVARIATE ANALYSIS



Terdapat outlier pada beberapa fitur seperti terlihat pada boxplot di atas.

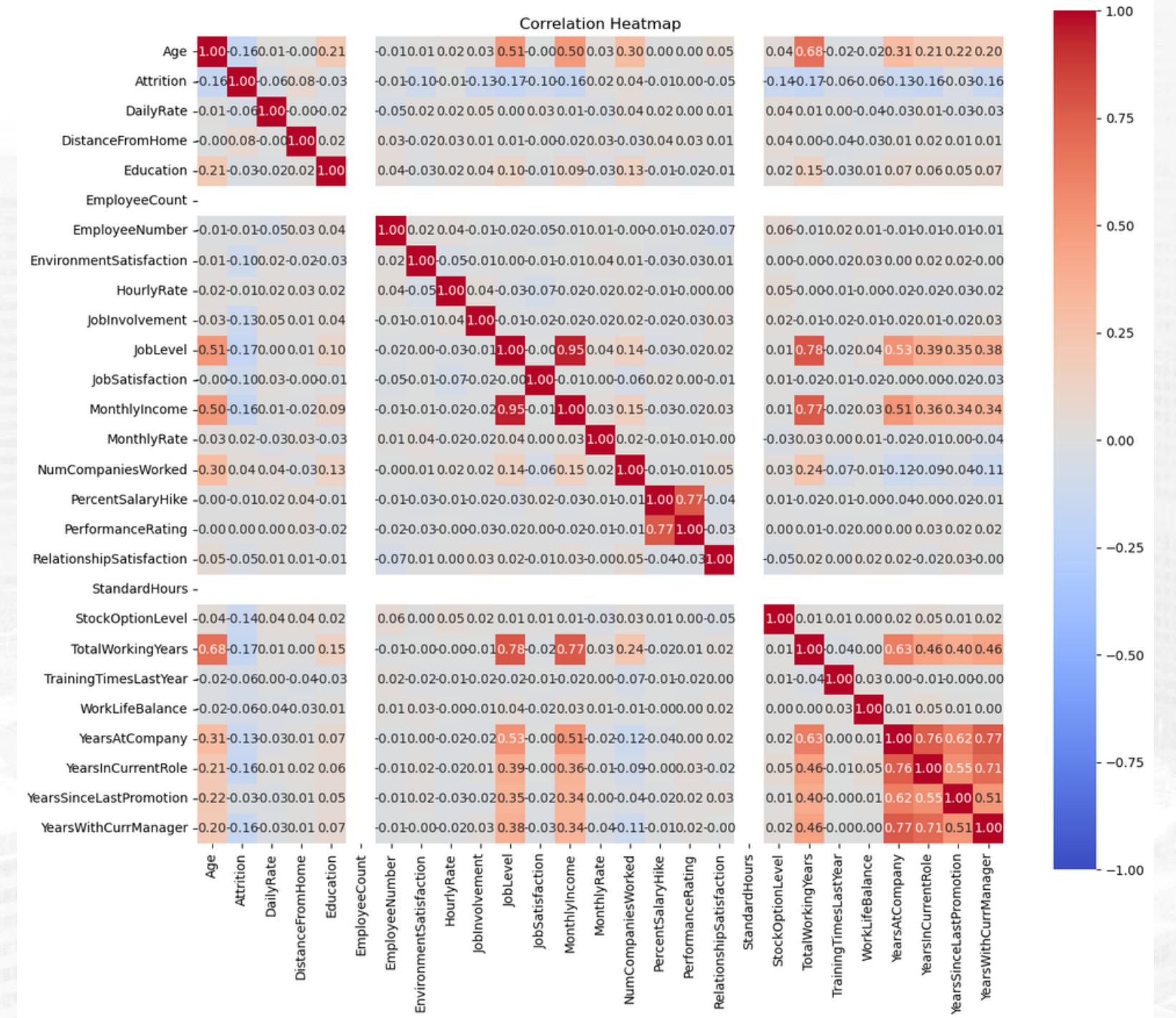
Rekomendasi:

- Perlu melakukan logtransformation pada fitur dengan jenis distribusi positively skewed
- Melakukan handling outlier dengan Z-score > 3 pada fitur-fitur yang memiliki outlier

#MULTIVARIATE ANALYSIS



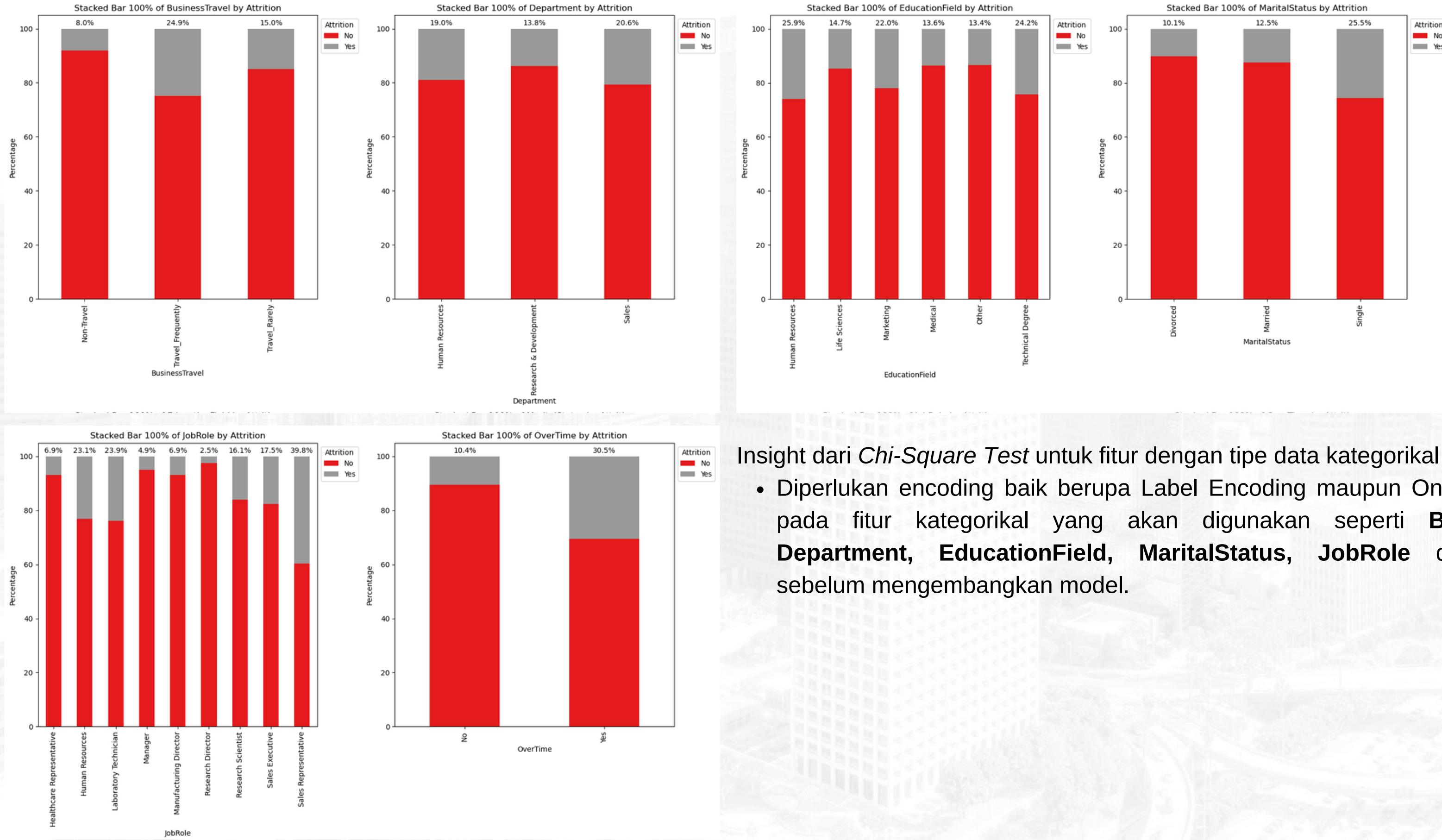
Pada tahap ini kami melakukan *feature selection* dengan *heatmap correlation table* untuk fitur dengan tipe data numerikal, dan *Chi-Square test* untuk fitur dengan tipe data kategorikal.



Insight dari *heatmap correlation table* :

- Fitur EmployeeCount dan StandardHours tidak relevan dan harus di-drop
- Fitur yang paling berhubungan dengan Attrition adalah JobLevel & TotalWorkingYears berkorelasi secara negatif atau bermakna semakin kecil value di JobLevel dan TotalWorkingYears maka semakin besar kecenderungan untuk Attrition.
- Hubungan antara fitur MonthlyIncome dengan JobLevel adalah fitur dengan korelasi tertinggi sebesar 0.95. Hal ini menandakan bahwa keduanya saling berkaitan dan mempengaruhi satu sama lain.
- Fitur YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion dan YearsWithCurrManager adalah kumpulan fitur yang memiliki korelasi yang tinggi dan saling berhubungan

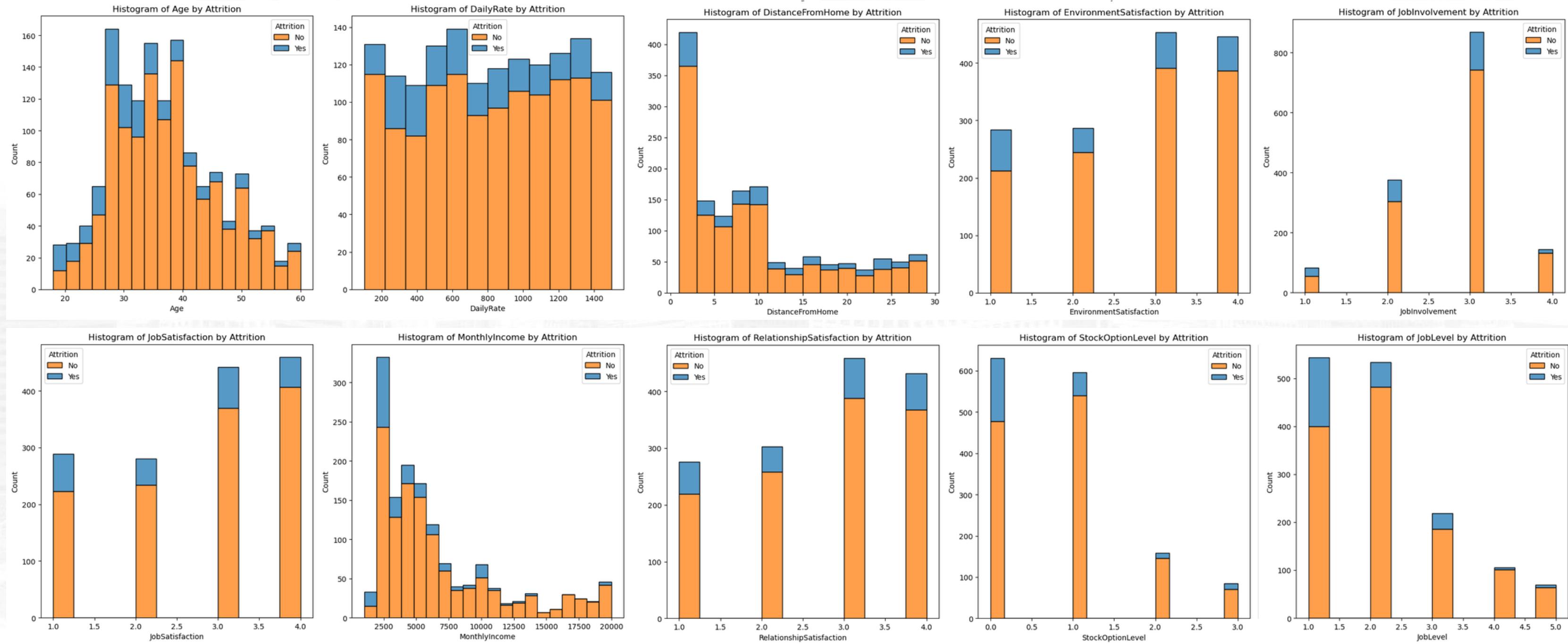
#MULTIVARIATE ANALYSIS



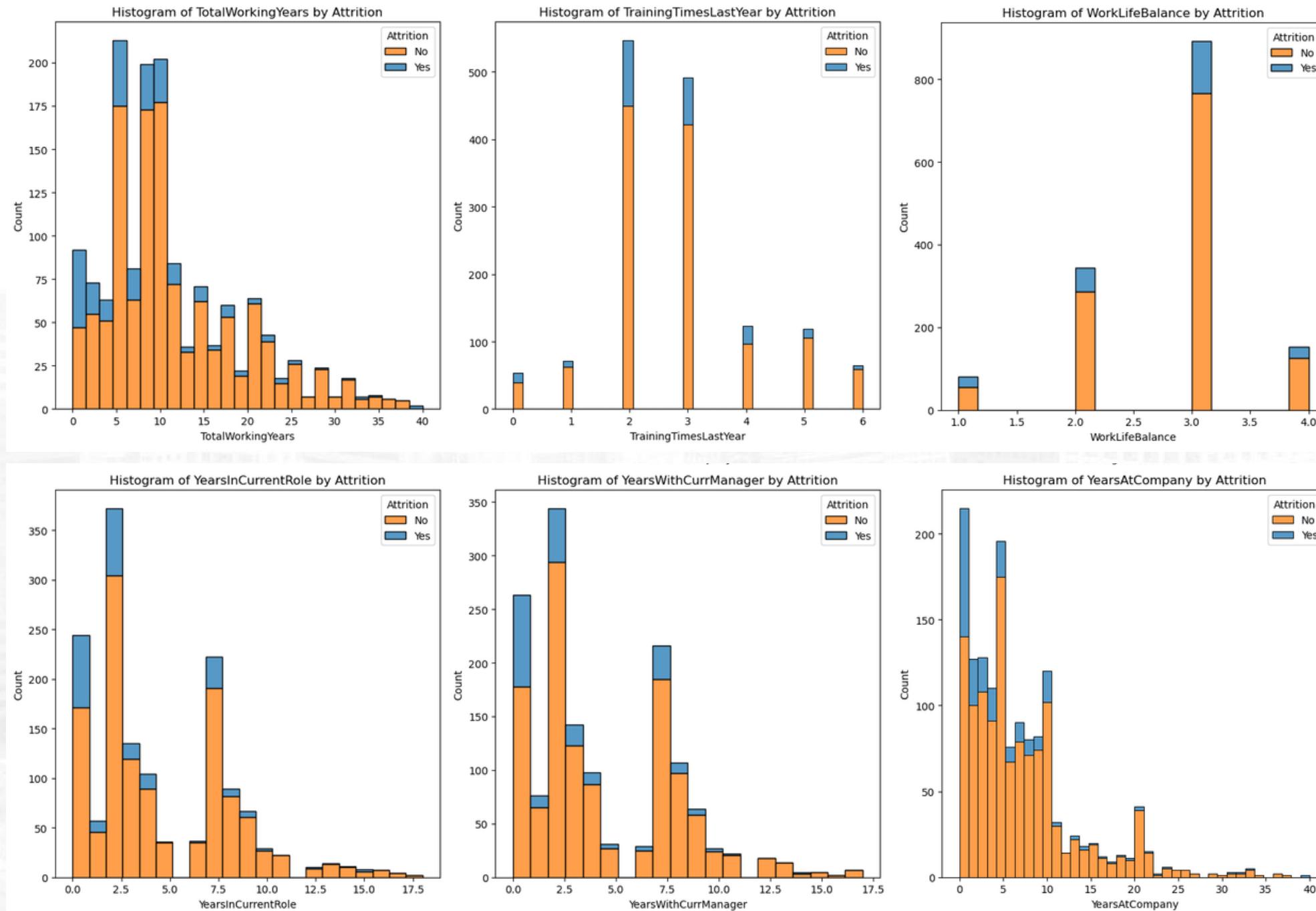
Insight dari *Chi-Square Test* untuk fitur dengan tipe data kategorikal :

- Diperlukan encoding baik berupa Label Encoding maupun One-Hot Encoding pada fitur kategorikal yang akan digunakan seperti **BusinessTravel**, **Department**, **EducationField**, **MaritalStatus**, **JobRole** dan **OverTime** sebelum mengembangkan model.

#MULTIVARIATE ANALYSIS



#MULTIVARIATE ANALYSIS



Insight untuk fitur dengan tipe data numerikal :

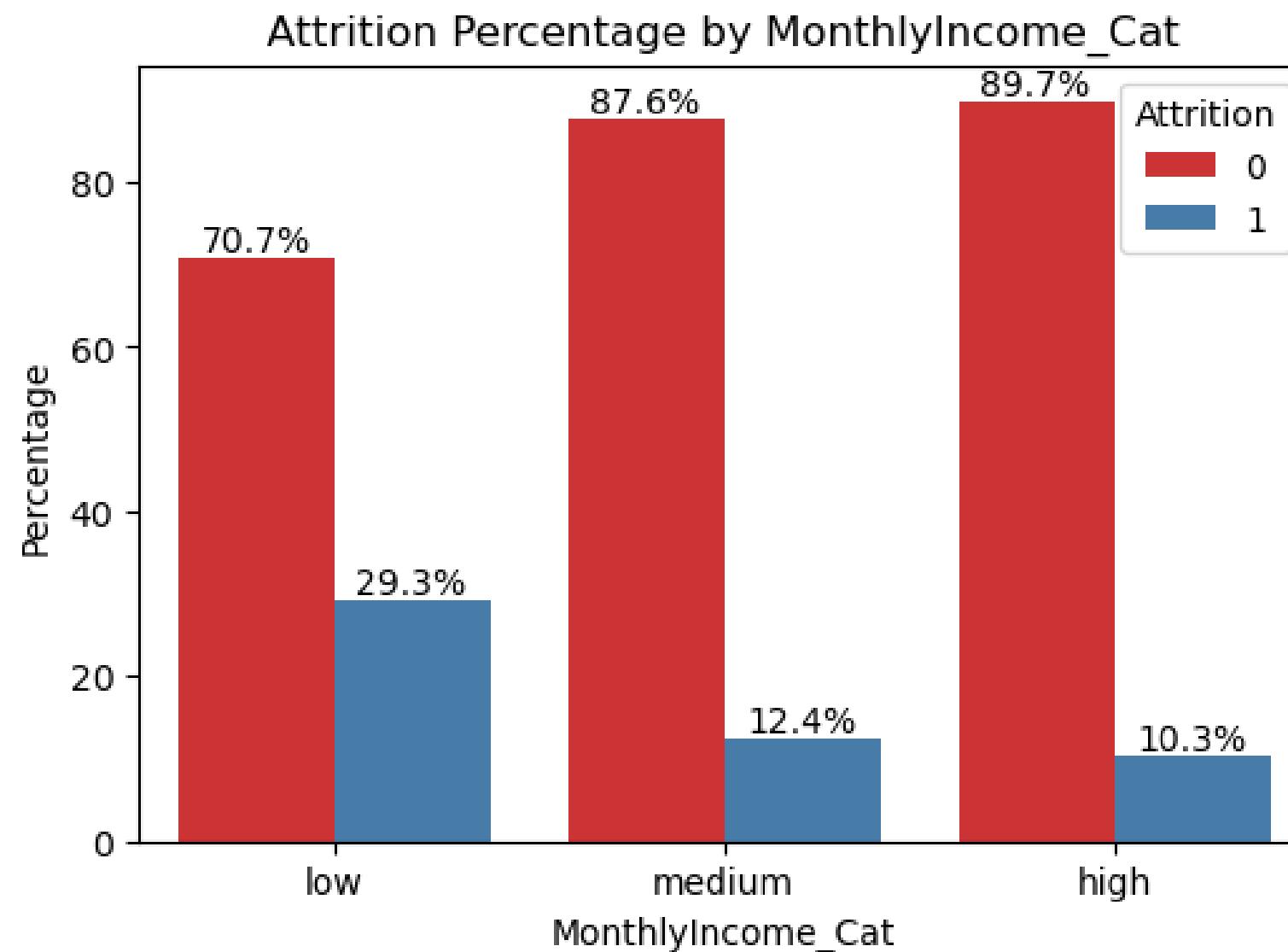
- Terlihat bahwa **StandardHours**, **EmployeeCount**, **EmployeeNumber** dan **Over18** hanya terdiri dari 1 jawaban sehingga tidak diperlukan sebagai fitur model.
- Berdasarkan corelation heatmap, fitur numeric yang dipertahankan dengan threshold 0.05 yaitu : **Age**, **DailyRate**, **DistanceFromHome**, **EnvironmentSatisfaction**, **JobInvolvement**, **MonthlyIncome**, **RelationshipSatisfaction**, **StockOptionLevel**, **TotalWorkingYears**, **TrainingTimeLastYar**, **WorkLifeBalance**, **YearsAtCompany**, **YearsICurrentRole**, **YearsWithCurrManager**

#Insights and Business Recommendations

##Analisis 1 (Membuat Kategori MonthlyIncome terhadap Attrition)



Pada tahap ini, kami memodifikasi fitur MonthlyIncome menjadi MonthlyIncome_Cat di mana nilai MonthlyIncome dikelompokkan menjadi 3 kategori Low, Medium dan High untuk mempermudah analisis. Hal ini didasarkan pada nilai kuartil 1, 2 dan 3 pada fitur MonthlyIncome.

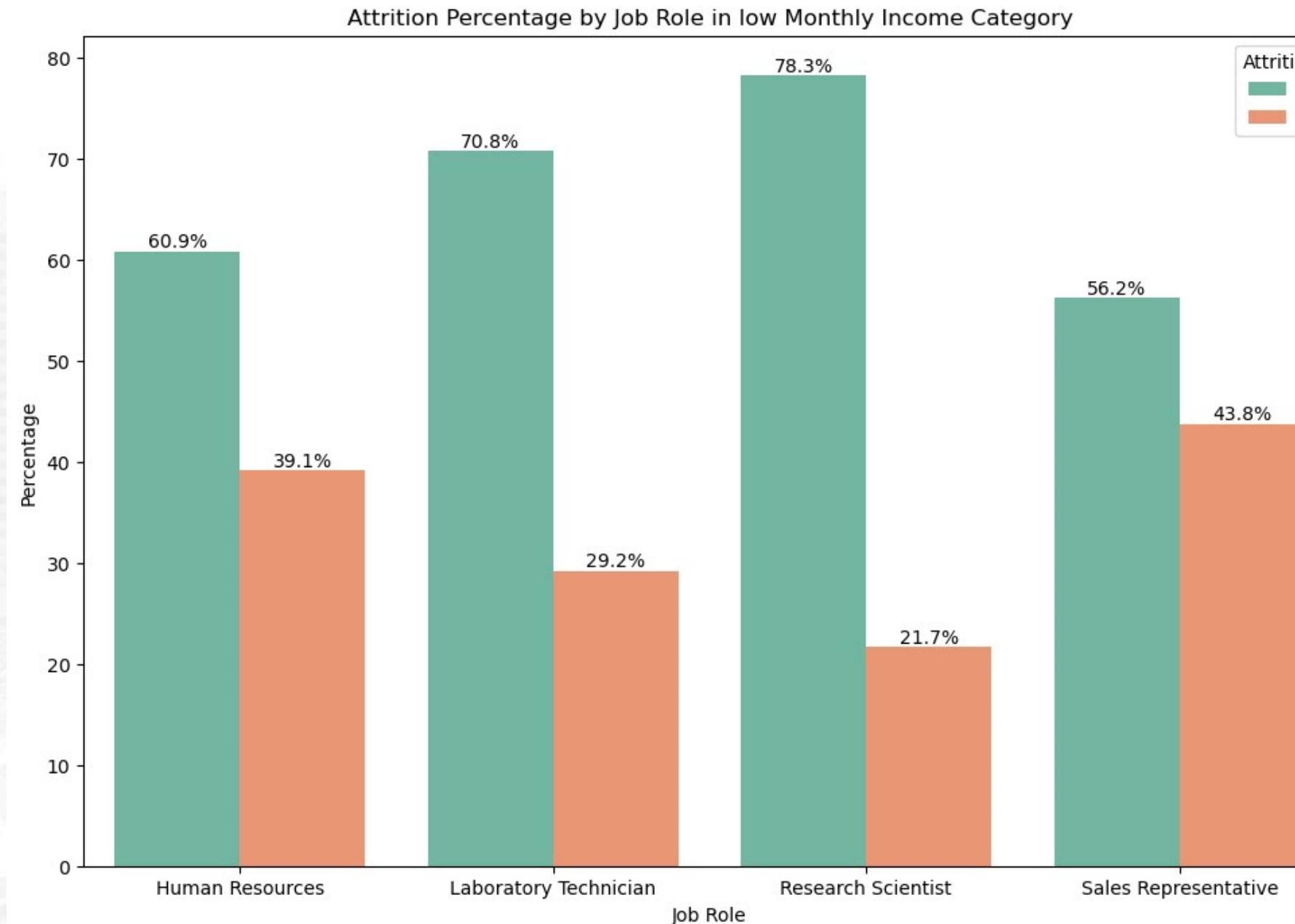


Insights :

- Berdasarkan visualisasi data MonthlyIncome_Cat vs Attrition menunjukkan bahwa hampir 30% rasio karyawan pada kategori Low Income adalah 'Yes' Attrition. Sementara pada karyawan yang tergolong kategori medium - high monthly income karyawan yang memiliki kecenderungan untuk churn/attrition hanya sebesar 10-12%. Hal ini mengindikasikan bahwa faktor low income mempengaruhi keputusan Attrition pada karyawan.

#Insights and Business Recommendations

##Analisis 2 (Melihat Persebaran Karyawan dengan MonthlyIncome Kategori 'low' pada JobRole terhadap Attrition)



Insights :

- Persebaran karyawan dengan kategori low MonthlyIncome berdasarkan JobRole menunjukkan bahwa jumlah karyawan dengan low MonthlyIncome adalah karyawan Laboratory Technician sebanyak 38 orang. Sementara secara rasio, karyawan Sales Representative adalah yang tertinggi yaitu sebesar 43.8% yang berisiko untuk Attrition.



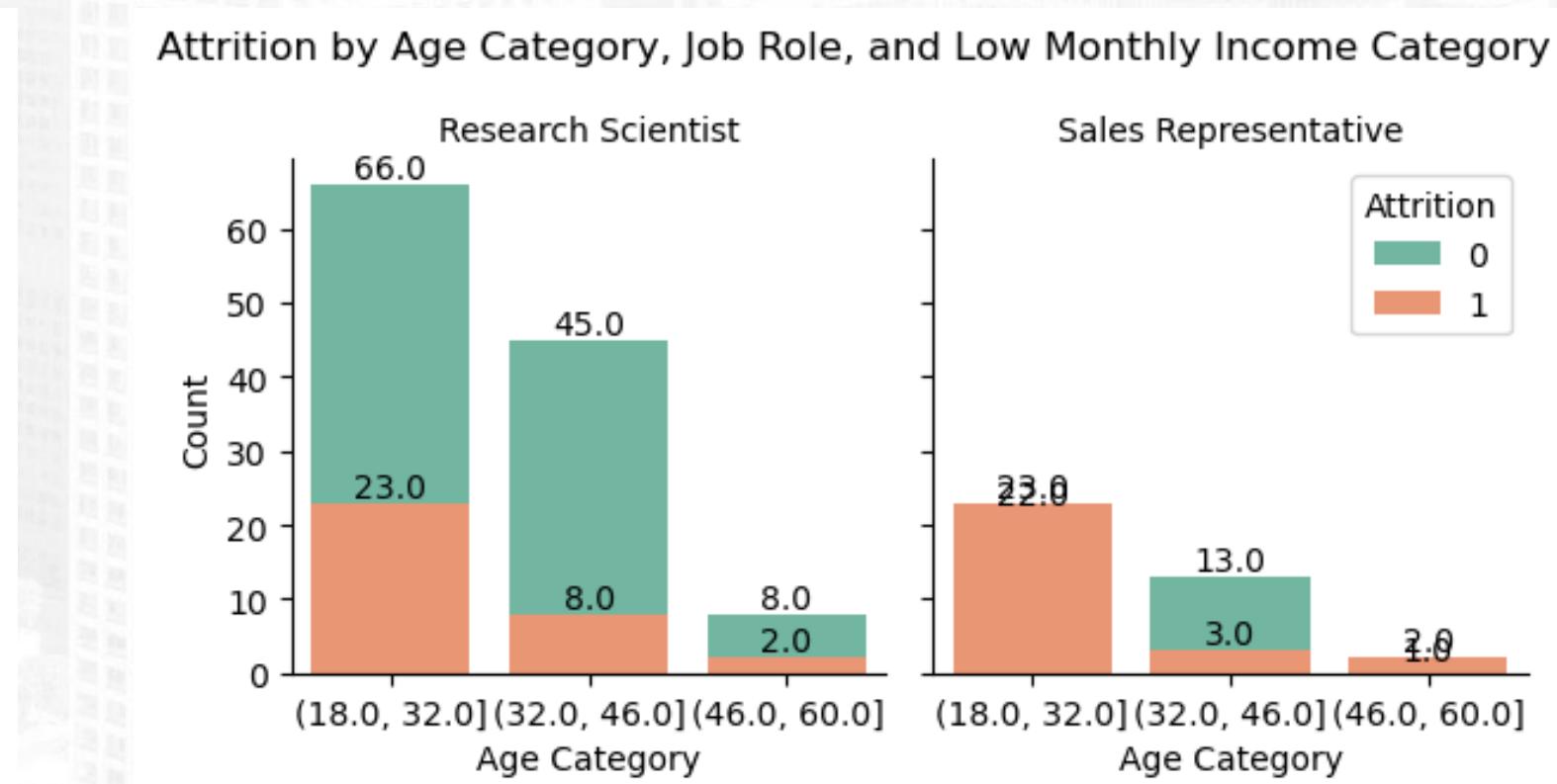
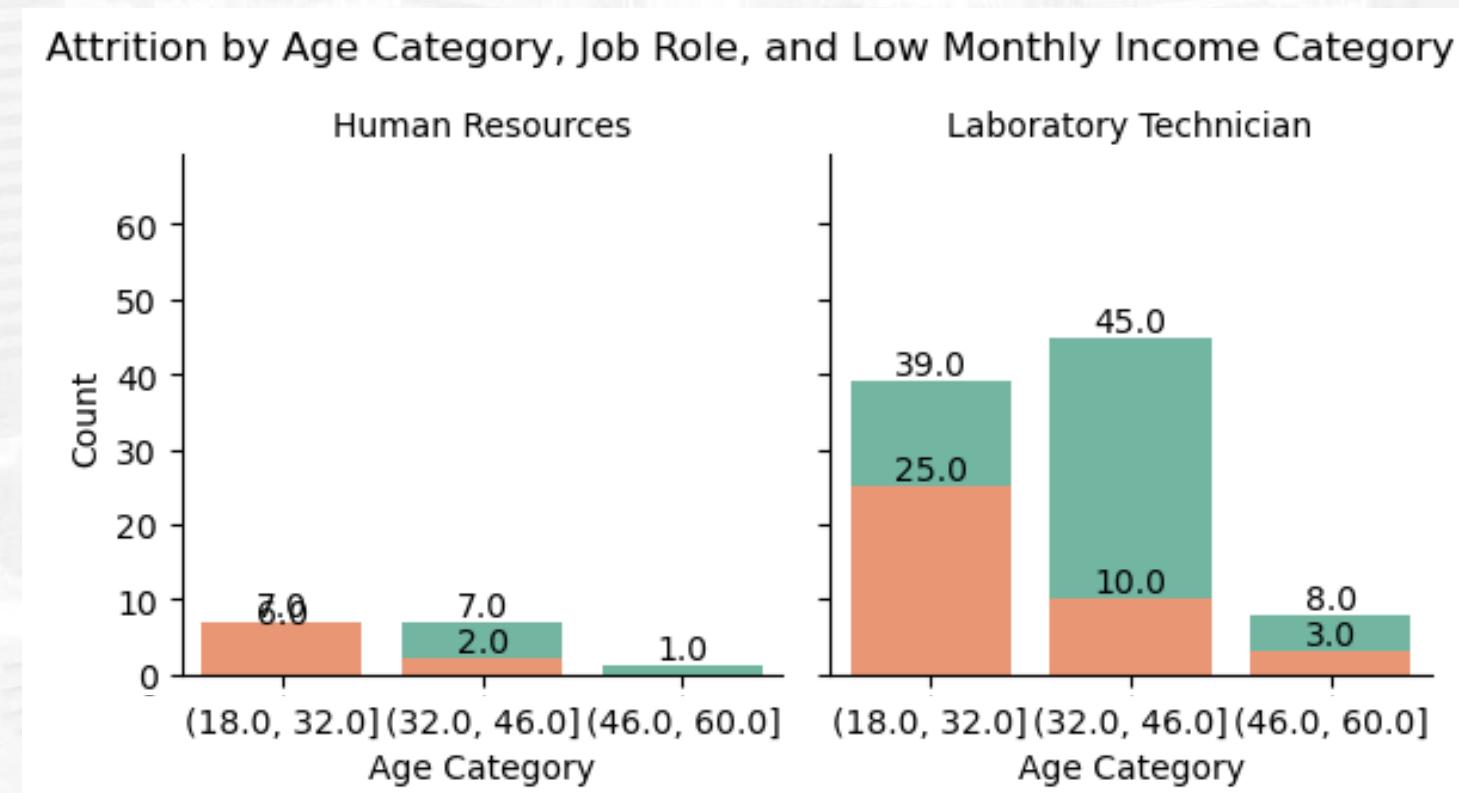
#Insights and Business Recommendations

##Analisis 3 (Fitur Age_Cat dihubungkan dengan JobRole, 'low' MonthlyIncome dan Attrition)

Pada tahap ini, kami memodifikasi fitur 'Age' menjadi 'Age_Cat' di mana nilai 'Age' dikelompokkan menjadi :

- 0 41 (32.0, 46.0]
- 1 49 (46.0, 60.0]
- 2 37 (32.0, 46.0]
- 3 33 (32.0, 46.0]
- 4 27 (18.0, 32.0]

Kemudian, kami melakukan filter data untuk kategori 'low' pada kolom 'MonthlyIncome_Cat' dan menghitung jumlah karyawan dalam setiap kombinasi 'Age_Cat', 'JobRole', dan 'Attrition' di kategori 'low' MonthlyIncome_Cat.



Insights :

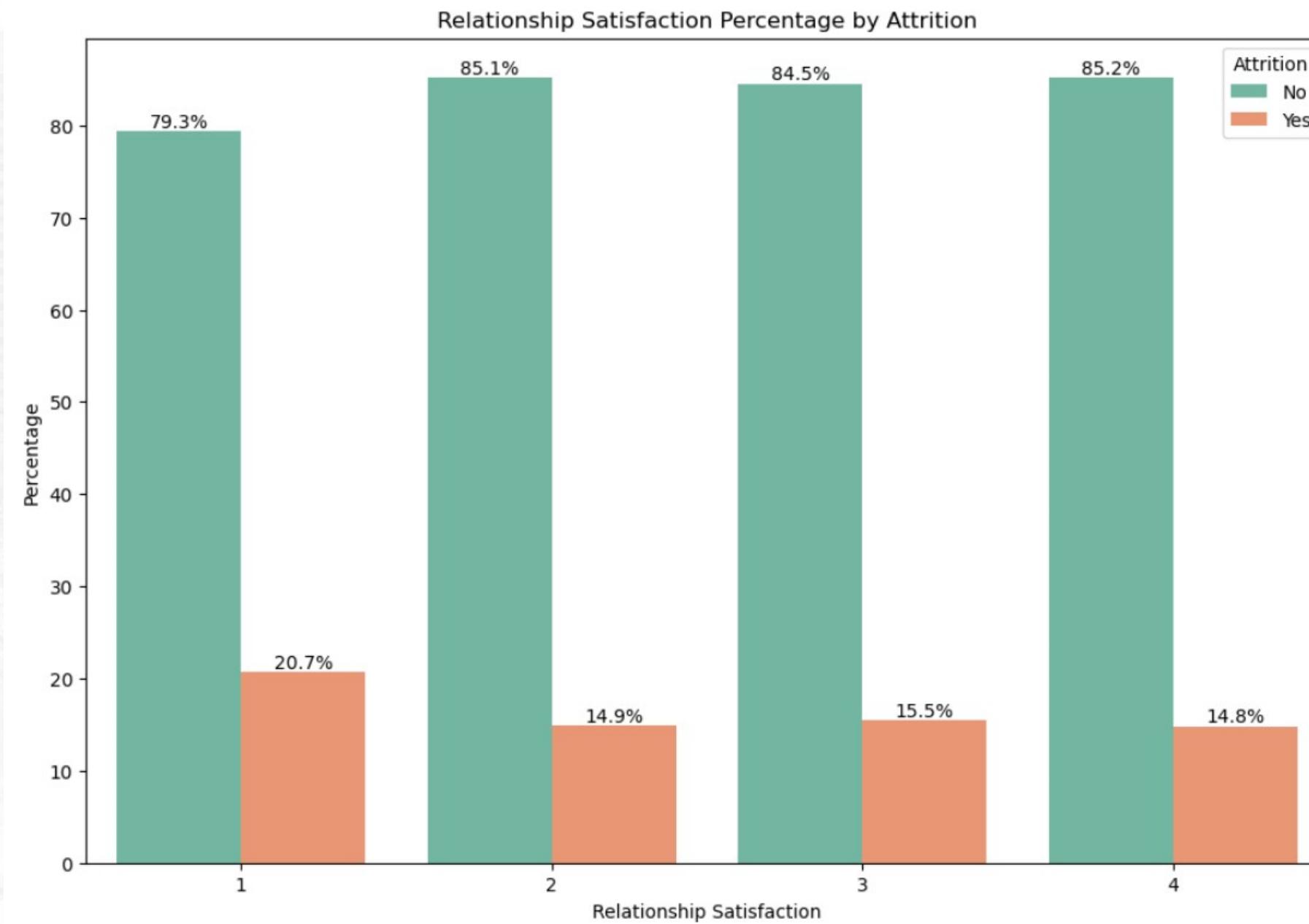
- Karyawan pada rentang usia 18-32 tahun cenderung yes attrition pada seluruh job role
- karyawan pada rentang usia 32-46 tahun cenderung no attrition pada seluruh job role
- Karyawan yang paling banyak yes attrition berasal dari departemen sales dengan rentang umur 18-32 tahun (100%), dan yang paling banyak no attrition berasal dari departemen research dengan rentang umur 18-32 tahun



#Insights and Business Recommendations

##Analisis 4 (Fitur Relationship Satisfaction dihubungkan dengan Attrition)

Pada feature RelationshipSatisfaction, kami membuat visualisasi antar value category dan membuat persentase perbandingan masing-masing kategori value terhadap feature attrition karyawan. Kategori terbagi atas 4, yaitu 1 yang didefinisikan sebagai kategori Low, 2 didefinisikan sebagai kategori Medium, 3 didefinisikan sebagai kategori High dan 4 didefinisikan sebagai kategori Very High.



Insights :

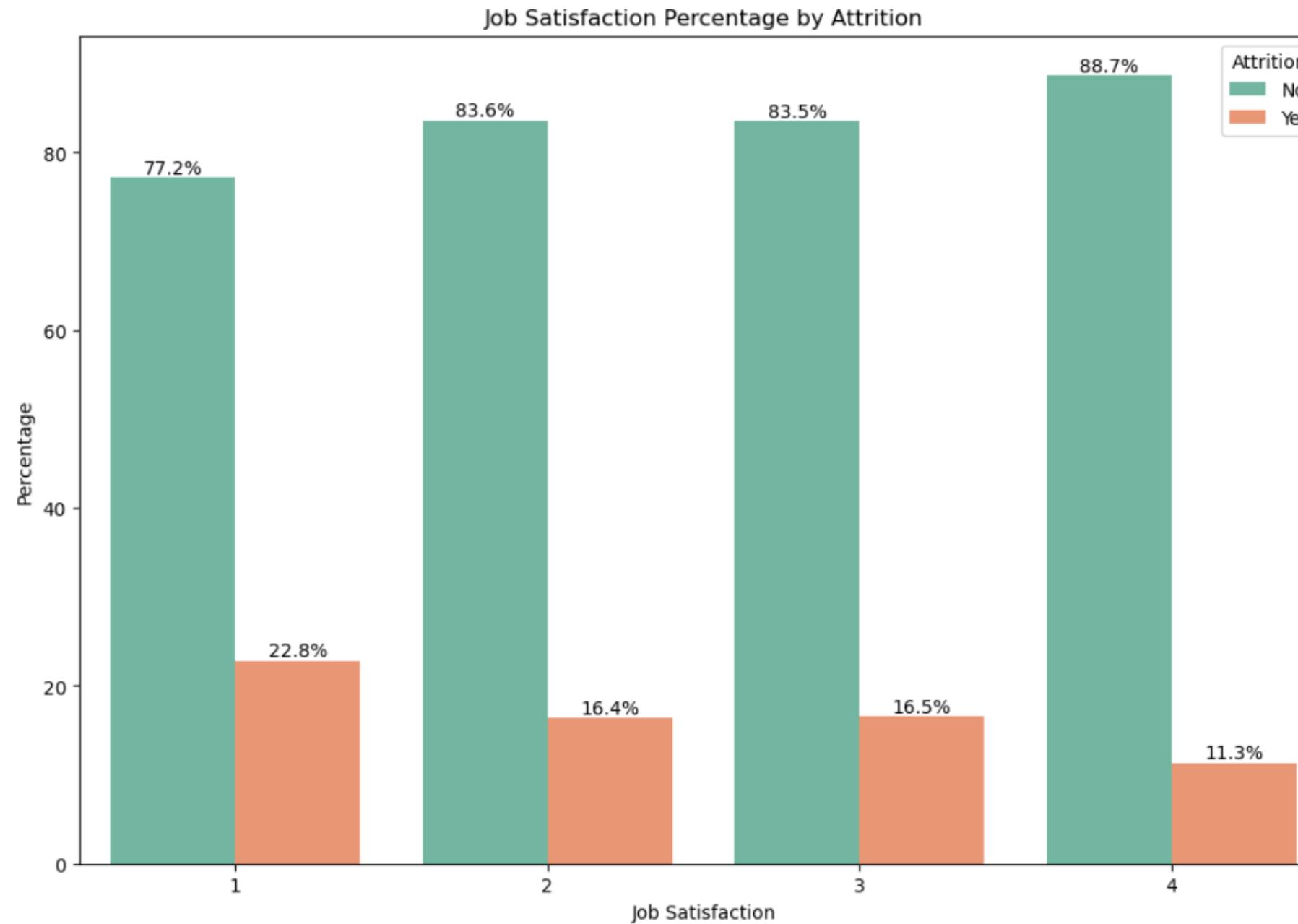
Berdasarkan visualisasi di atas, diketahui bahwa karyawan pada kategori 1 yang terdefenisi sebagai Low Satisfied memiliki persentasi 'Yes' Attrition tertinggi dari kategori lainnya yaitu sebesar 20.7%. Hal ini berarti tingkat kepuasan karyawan yang rendah terhadap hubungan maupun lingkungan kerja diperusahaan mereka cukup mempengaruhi keputusan karyawan untuk memilih 'Yes' Attrition pada kategori ini meskipun persentasi terhadap karyawan yang memilih 'No' Attrition jauh lebih dominan.

#Insights and Business Recommendations

##Analisis 5 (Fitur Job Satisfaction dihubungkan dengan Attrition)



Pada feature JobSatisfaction, kami membuat visualisasi antar value category dan membuat persentase perbandingan masing-masing kategori value terhadap feature attrition karyawan. Kategori terbagi atas 4, yaitu 1 yang didefinisikan sebagai kategori Low, 2 didefinisikan sebagai kategori Medium, 3 didefinisikan sebagai kategori High dan 4 didefinisikan sebagai kategori Very High.



Insights :

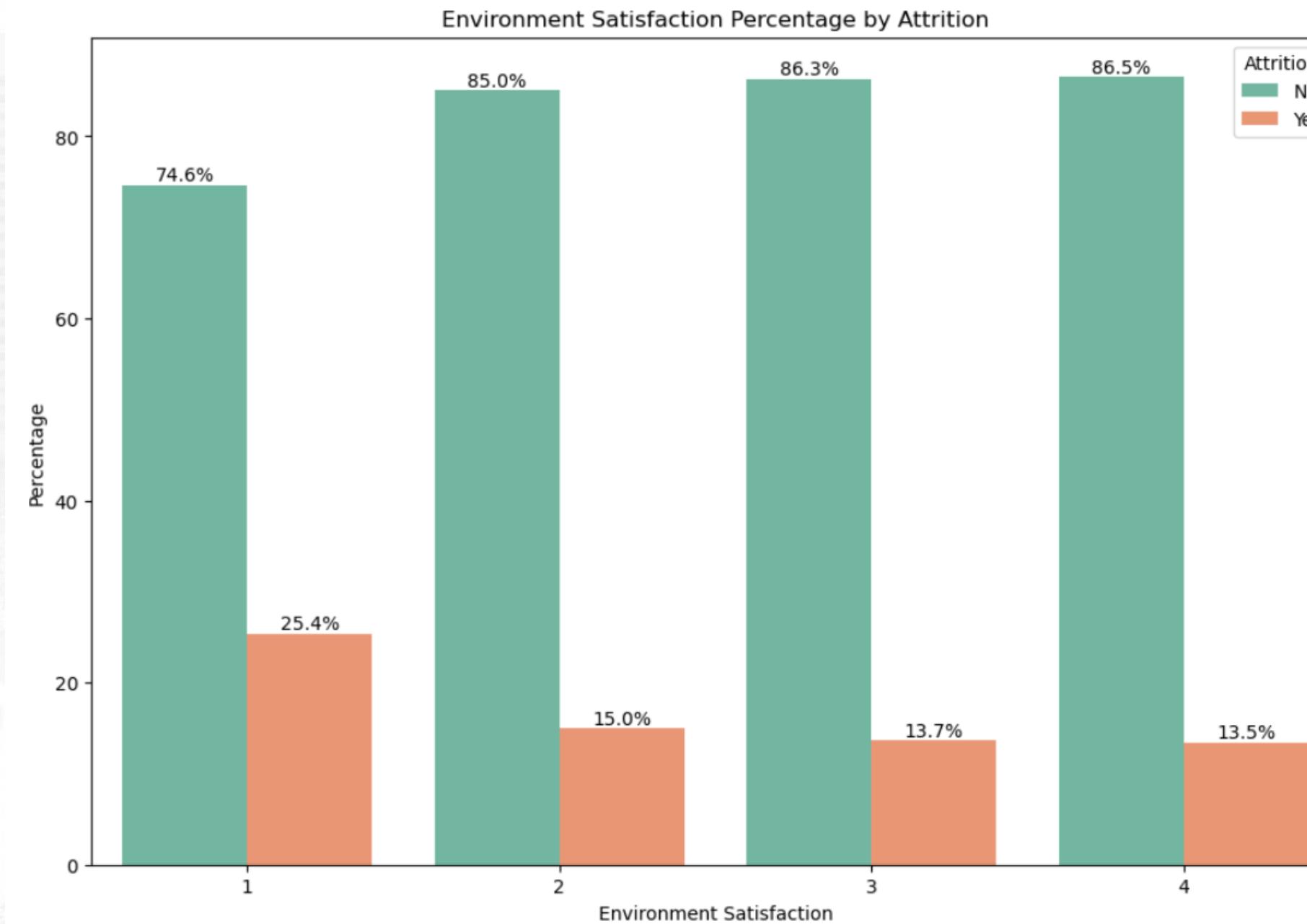
Berdasarkan visualisasi di atas, diketahui bahwa karyawan pada kategori 1 yang terdefenisi sebagai Low Satisfied memiliki persentasi 'Yes' Attrition tertinggi dari kategori lainnya yaitu sebesar 22.8%. Hal ini berarti tingkat kepuasan karyawan yang rendah terhadap hubungan maupun lingkungan kerja diperusahaan mereka cukup mempengaruhi keputusan karyawan untuk memilih 'Yes' Attrition pada kategori ini meskipun persentasi terhadap karyawan yang memilih 'No' Attrition jauh lebih dominan.

#Insights and Business Recommendations

##Analisis 6 (Fitur Environment Satisfaction dihubungkan dengan Attrition)



Pada feature JobSatisfaction, kami membuat visualisasi antar value category dan membuat persentase perbandingan masing-masing kategori value terhadap feature attrition karyawan. Kategori terbagi atas 4, yaitu 1 yang didefinisikan sebagai kategori Low, 2 didefinisikan sebagai kategori Medium, 3 didefinisikan sebagai kategori High dan 4 didefinisikan sebagai kategori Very High.



Insights :

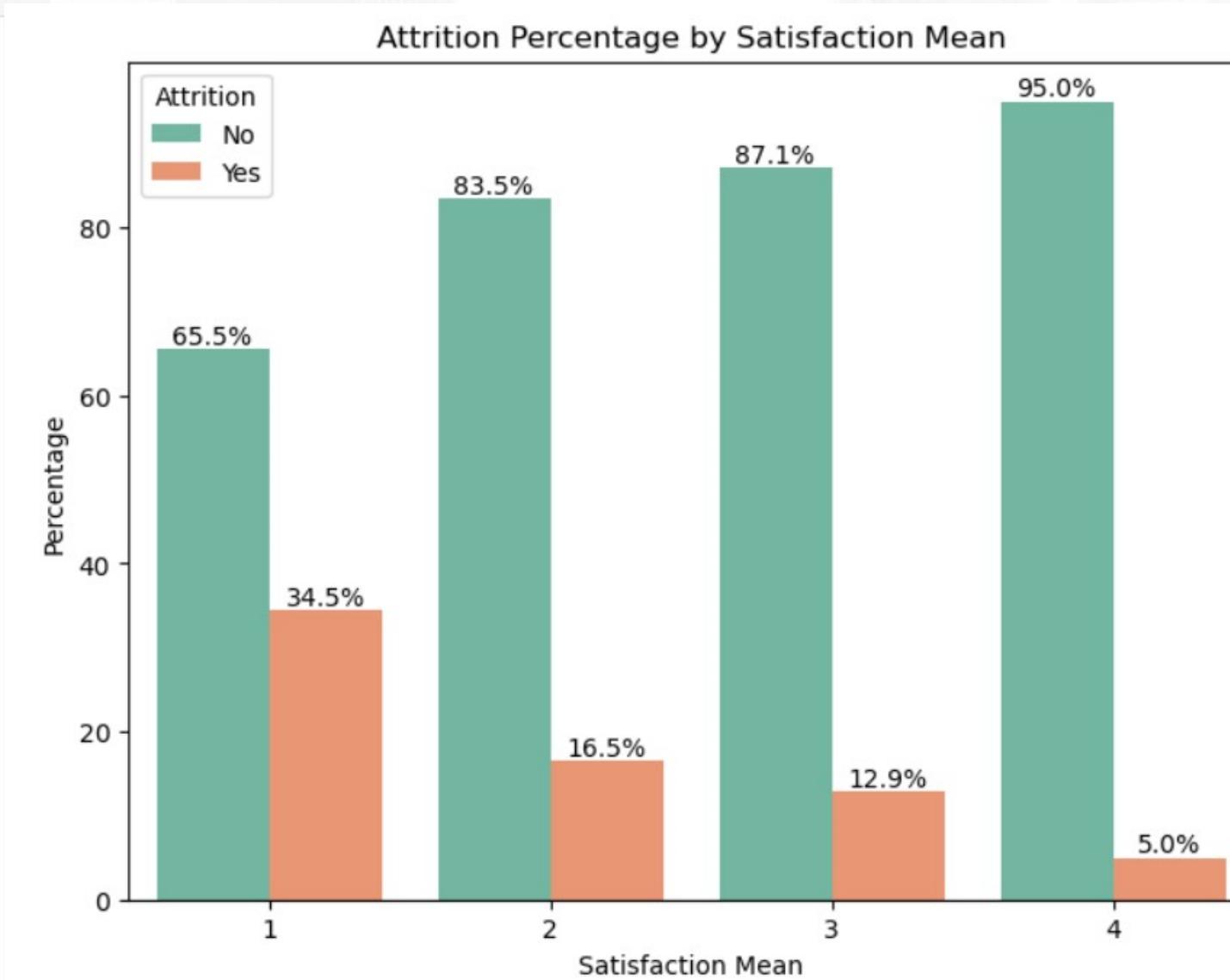
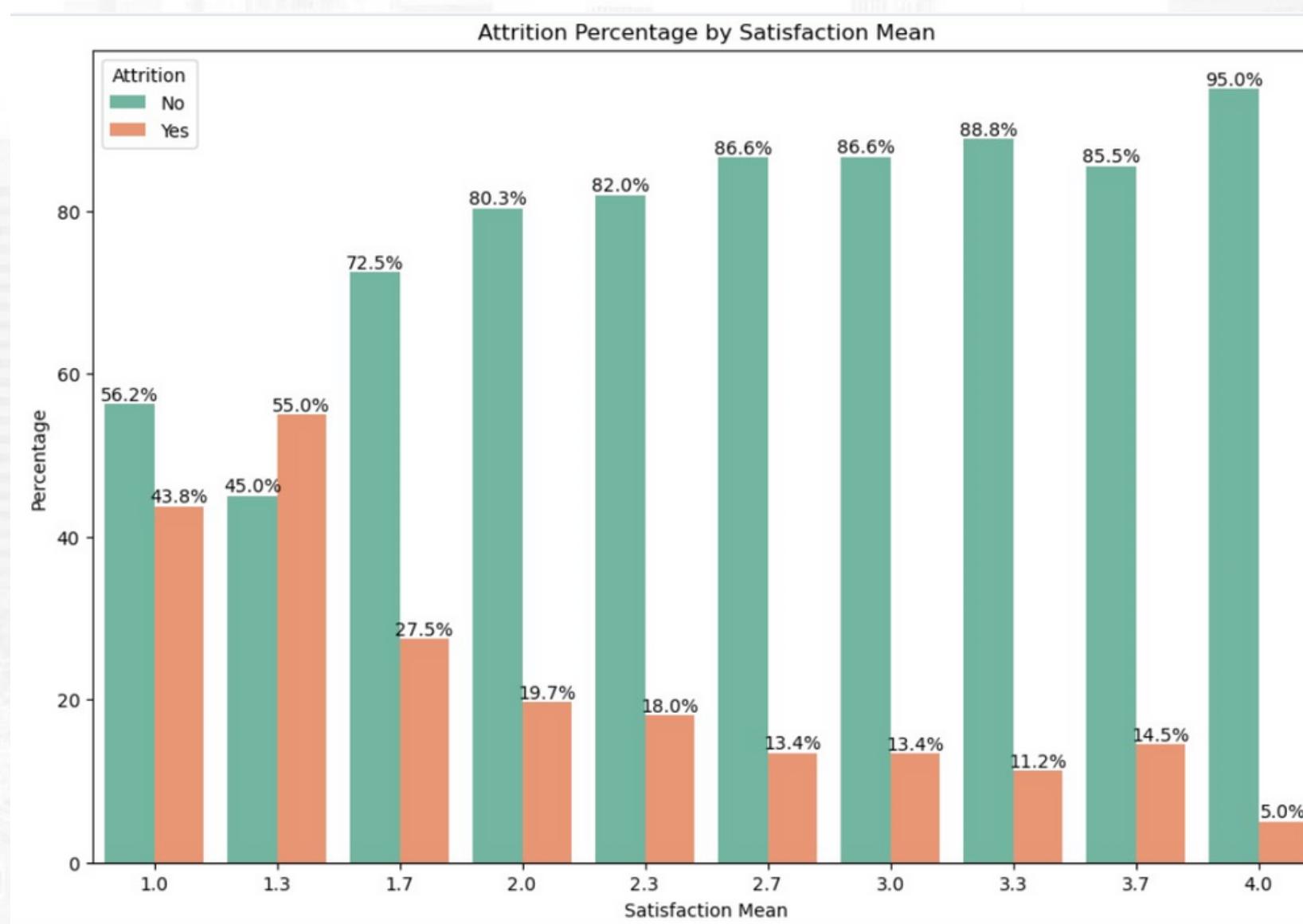
Berdasarkan visualisasi di atas, diketahui bahwa karyawan pada kategori 1 yang terdefenisi sebagai Low Satisfied memiliki persentasi 'Yes' Attrition tertinggi dari kategori lainnya yaitu sebesar 25.4%. Hal ini berarti tingkat kepuasan karyawan yang rendah terhadap hubungan maupun lingkungan kerja diperusahaan mereka cukup mempengaruhi keputusan karyawan untuk memilih 'Yes' Attrition pada kategori ini meskipun persentasi terhadap karyawan yang memilih 'No' Attrition jauh lebih dominan.

#Insights and Business Recommendations

##Feature Satisfaction Mean terhadap Attrition



Kami membuat feature baru yaitu Satisfaction_Mean yaitu rata-rata dari setiap feature satisfaction diatas (Relationship, Job, dan Environment Satisfaction). Setelah itu kami mengkategorikan kembali feature baru tersebut menjadi rentang yang sama seperti pada feature satisfaction lainnya yaitu 1 - 4. kami membuat visualisasi antar value category dan membuat persentase perbandingan masing-masing kategori value terhadap feature attrition karyawan.



Insights :

Berdasarkan visualisasi di atas, diketahui bahwa semakin rendah Satisfaction mean (yang terdiri dari job satisfaction, relationship satisfaction, dan environment satisfaction), semakin tinggi kecenderungan karyawan untuk 'yes' attrition. Sebaliknya, kecenderungan karyawan untuk 'no' attrition semakin tinggi seiring dengan penambahan satisfaction mean. Sehingga disini satisfaction berpengaruh terhadap attrition karyawan.

#Insights and Business Recommendations

##Recommendation Analisis 4,5,6 (Feature Satisfaction Mean terhadap Attrition)



- **Survei Kepuasan Karyawan yang Mendalam:** Lakukan survei kepuasan karyawan yang lebih mendalam dan komprehensif. Survei ini harus dilakukan secara rutin untuk memahami dengan lebih baik perasaan dan pandangan karyawan terhadap perusahaan serta mendeteksi perubahan tren dan masalah yang mungkin muncul sehingga perusahaan dapat mengambil tindakan yang sesuai.
- **Peningkatan lingkungan kerja baik fisik dan non-fisik:** Lingkungan kerja yang ideal adalah lingkungan kerja yang mampu membuat karyawan produktif tanpa menurunkan motivasi kerja dan meningkatkan stress pekerja. Melalui penambahan fasilitas fisik dan kegiatan bonding yang baik tidak hanya meningkatkan kepuasan lingkungan kerja juga terhadap kepuasan kerja secara keseluruhan.
- **Beri umpan balik kepada kinerja karyawan dan ciptakan budaya engagement:** Umpan balik kinerja yang berkelanjutan memungkinkan karyawan untuk lebih mengetahui di mana posisi mereka, memberikan sarana formal untuk mendapat masukan, dan menunjukkan bahwa atasan memperhatikan mereka dan bahwa kinerja mereka penting. Hal ini guna meningkatkan hubungan dan memberikan lingkungan kerja yang lebih kondusif kepada karyawan. Karyawan yang bertanggung jawab dan dilibatkan dalam perusahaan akan lebih besar kemungkinannya untuk bertahan dengan atasan. Jadi, Manajer HR / atasan harus punya kebijakan moral untuk memberi reward kepada mereka yang loyal. Reward dalam konteks ini lebih dari sekedar kompensasi (gaji), namun juga mencakup dukungan kepada karyawan untuk mencapai tujuan pribadi yang selaras dengan tujuan perusahaan.
- **Fleksibel dalam hal work-life balance:** Pekerja sangat-sangat menghargai keseimbangan antara pekerjaan dan kehidupan pribadi. Mereka ingin cara yang fleksibel untuk dapat terlibat dengan atasannya. Untuk menarik dan mempertahankan pekerja dengan pengalaman dan ekspektasi karir yang berbeda, perusahaan harus lebih fleksibel dalam menstrukturisasi beban pekerjaan dan ekspektasi karyawan. Hal ini berkaitan dengan upaya memenuhi ekspektasi hubungan dan lingkungan kerja di perusahaan.