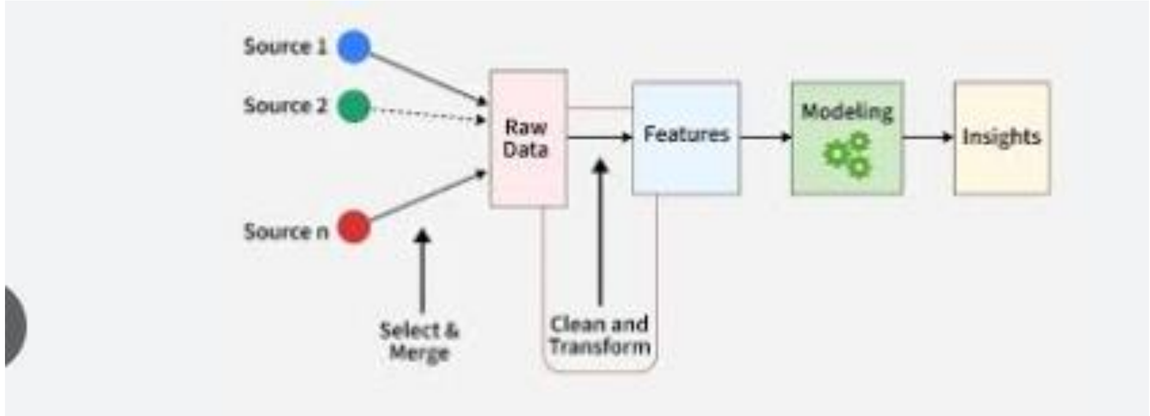
Customer-Churn _ **Feature Engineering & Data Transformation**(Week_4)

DATA SCIENCE INTERNSHIP -FUTUREXCEL

ANALYTIC BY: **MUHAMMAD ALI ABDUL WAHID**

Introduction

- This task focuses on feature engineering for the Customer Churn dataset. The aim was to transform existing customer information into meaningful features that help in understanding customer behavior and retention patterns.



Dataset Used

- The cleaned Customer Churn dataset from Week 1 was used as input for this task. The dataset was already prepared and suitable for feature transformation

Load Data

```
import pandas as pd
df = pd.read_csv("/content/cleaned_customer_churn_data (3).csv")
df.head()
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	Dev
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	
4	9237-	Female	0	No	No	2	Yes	No	Fiber optic	No	...	

Feature Engineering

- New features such as customer tenure groups, charge levels, and churn indicators were created. These features help simplify complex data and make customer behavior easier to analyze.

Feature Engineering

```
df['tenure_group'] = df['tenure'].apply(
    lambda x: 'Short' if x<12 else 'Medium' if x<36 else 'Long'
)

df['monthly_charge_level'] = df['MonthlyCharges'].apply(
    lambda x: 'Low' if x<35 else 'Medium' if x<70 else 'High'
)

df['is_long_term'] = df['tenure'].apply(lambda x: 1 if x>=24 else 0)

df['churn_flag'] = df['Churn'].apply(lambda x: 1 if x=='Yes' else 0)
df.head()
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	Con
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	Mo
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	On
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	Mo
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	On
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	Mo

5 rows × 25 columns

Encoding and Scaling

- Categorical features were encoded into numerical values, while important numerical features were scaled. This ensures uniformity and improves the dataset's readiness for future modeling tasks.

Encoding

```python
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
df['tenure_group_encoded'] = le.fit_transform(df['tenure_group'])
df.head()
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	Pap
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	

Scaling

```python
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
df['MonthlyCharges_scaled'] = scaler.fit_transform(df[['MonthlyCharges']])
df
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...

Before and After Comparison

- After transformation, the dataset shows better numerical consistency and improved feature representation compared to the original version.

Before vs After

```
df[['MonthlyCharges','MonthlyCharges_scaled']].describe()
```

	MonthlyCharges	MonthlyCharges_scaled
count	7043.000000	7.043000e+03
mean	64.761692	-6.406285e-17
std	30.090047	1.000071e+00
min	18.250000	-1.545860e+00
25%	35.500000	-9.725399e-01
50%	70.350000	1.857327e-01
75%	89.850000	8.338335e-01
max	118.750000	1.794352e+00

Conclusion

- Through this task, I gained practical experience in feature engineering and learned how transformed features can support better decision-making and predictive analysis.