

DATA SCIENCE INTERNSHIP -  
FUTUREXCEL

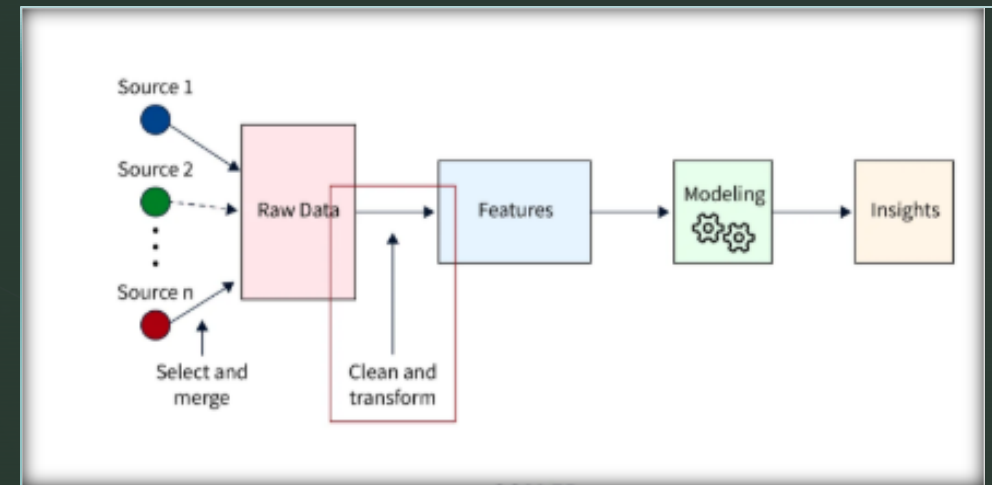
ANALYTIC BY :REHANA ABDUL.WAHID

# ▼ AIRLINE -PASSENGER- DATASET – Feature Engineering & Data Transformation(Week\_4)



# Introduction

- In this task, feature engineering was performed on the Airline Passenger Satisfaction dataset. The goal was to create new features that capture passenger experience and service quality more effectively.



# Dataset Used

```
import pandas as pd
df = pd.read_csv("/content/cleaned_airline_satisfaction (1) (1).csv")
df.head()
```

	Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	...	Inflight entertainment	On-board service	Leg room service	Baggage handling	Check-in service
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	...	5	4	3	4	
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	2	...	1	1	5	3	
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2	...	5	4	3	4	
3	3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	5	...	2	2	5	3	
4	4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	3	...	3	3	4	4	

- The cleaned airline dataset prepared earlier was used for feature engineering. This allowed smooth transformation without additional preprocessing steps.

# Feature Engineering

## Feature Engineering

```
df['total_service_score'] = (
    df['Seat comfort'] +
    df['Inflight service'] +
    df['Cleanliness']
)

df['is_loyal_customer'] = df['Customer Type'].apply(
    lambda x: 1 if x=='Loyal Customer' else 0
)

df['delay_flag'] = df['Arrival Delay in Minutes'].apply(
    lambda x: 1 if x>15 else 0
)
```

...	Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	...	Baggage handling	Checkin service	Inflight service	Cleanliness
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	...	4	4	5	5
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	2	...	3	1	4	1
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2	...	4	4	4	5
3	3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	5	...	3	1	4	2
4	4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	3	...	4	3	3	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
101608	103899	94171	Female	disloyal Customer	23	Business travel	Eco	192	2	1	...	4	2	3	2
101609	103900	73097	Male	Loyal Customer	49	Business travel	Business	2347	4	4	...	5	5	5	4

New features were derived by combining service-related scores, identifying loyal customers, and categorizing flight-related information. These features provide a clearer view of passenger satisfaction patterns.

# Encoding and Scaling

- Categorical variables were encoded into numerical form, and numerical features were scaled to ensure consistency across the dataset. These steps help standardize the data.

## Encoding

```
from sklearn.preprocessing import LabelEncoder  
  
le = LabelEncoder()  
  
df['satisfaction_encoded'] = le.fit_transform(df['satisfaction'])  
  
df[['satisfaction', 'satisfaction_encoded']].head()
```

	satisfaction	satisfaction_encoded
0	neutral or dissatisfied	0
1	neutral or dissatisfied	0
2	satisfied	1
3	neutral or dissatisfied	0
4	satisfied	1


## Scaling

```
[14]  
✓ Os  
from sklearn.preprocessing import StandardScaler  
scaler = StandardScaler()  
df['Flight_Distance_scaled'] = scaler.fit_transform(df[['Flight Distance']])
```

# Before and After Comparison

- The comparison between the original and transformed dataset shows improved feature structure and normalized values, making the dataset suitable for further analysis.

Before vs After

[11] ✓ Os  `df[['Flight Distance', 'Flight_Distance_scaled']].describe()`

	Flight Distance	Flight_Distance_scaled
count	101613.000000	1.016130e+05
mean	1128.632537	-3.020819e-17
std	921.114146	1.000005e+00
min	31.000000	-1.191642e+00
25%	407.000000	-7.834383e-01
50%	814.000000	-3.415799e-01
75%	1660.000000	5.768776e-01
max	3736.000000	2.830681e+00



# Conclusion

- This task improved my understanding of how feature engineering enhances raw datasets and prepares them for advanced analytical and machine learning applications.

