Data Cleaning Steps in Excel

Data Import

Imported three datasets:

- Flights dataset (5.7 million records)
- Airlines dataset
- Airports dataset

Header Standardization

- Promoted first row as headers.

Data Type Correction

- Year, Month, Day → Whole Number
- Airline → Text
- Delay columns → Whole Number
- Cancelled → Whole Number (0/1)
- Flight Date → Date
- Removed data type errors.

Created Standard Date Column

- Combined Year, Month, and Day into one column
- Converted it to Date format.

Handling Missing Values

- Identified null values in delay-related columns.
- Replaced null values in numeric delay columns with 0.

Created Derived Column – Flight Status

- Created new column using conditional logic
  If Cancelled = 1 → "Cancelled"
  If Arrival Delay > 15 → "Delayed"
  Else → "On Time"

Removed Unnecessary Columns

- Removed columns that were not required for analysis:
- Tail Number
- Taxi In
- Taxi Out

- Wheels On
- Wheels Off

Error Handling

- Removed rows containing errors.

Data Optimization

- Loaded cleaned dataset into Excel Data Model instead of worksheet.

Visualization Summary

Home

- The Home sheet provides a consolidated overview of airline performance by presenting key performance indicators such as Total Flights, Total Delayed Flights, Average Arrival Delay, and Average Departure Delay.
- A Gauge visualization is used to represent the overall On-Time performance percentage.
- A Donut chart illustrates the distribution of flights based on status (On-Time, Delayed, and Cancelled).
- Navigation buttons are implemented to allow to move the Delay Analysis and Cancellation Analysis sheets.

Delay Analysis

- The Delay Analysis sheet presents key delay-related KPIs including Average Arrival Delay, Total Weather Delay, Total Late Aircraft Delay, Total Security Delay, and Total Air System Delay.
- A Line Chart titled “Monthly Average Arrival Delay” is used to analyze delay trends across months.
- A Stacked Column Chart labeled “Delay Reason Breakdown” compares multiple delay categories (Weather, Security, Air System, and Late Aircraft) on a monthly basis.
- A “Top 10 Airlines by Average Delay” horizontal bar chart highlights airlines with the highest average arrival delay, enabling comparative performance analysis among carriers.
- A Gauge visualization displays the Average Arrival Delay.

Cancellation Analysis

- The Cancellation Analysis sheet presents key performance indicators including Total Cancelled Flights , Cancellation Rate , Most Common Cancellation Reason and the Airline with Highest Cancellation .
- A Column Chart titled “Total Cancelled Flights by Cancellation Reason” compares cancellation categories such as Weather, Airline, National Air System, and Security.
- A Bar Chart labeled “Total Cancelled Flights by Airline” highlights airlines with the highest number of cancellations, allowing comparative performance evaluation across carriers.
- A Line Chart titled “Monthly Cancellation Trend” displays cancellation fluctuations throughout the year, helping identify seasonal spikes and irregular patterns.
- A Donut Chart and Gauge visualization together illustrate cancellation distribution by reason and measure the cancellation rate against a defined performance benchmark.

Drill Through

- The Drill-Through page provides a detailed airline-level performance analysis by dynamically filtering data based on the selected airline from the Delay Analysis page**.**

Fact and Dimension Tables

Fact Table:

- Arrival Delay
- Departure Delay
- Weather Delay
- Air System Delay
- Security Delay
- Cancelled
- Date
- Airline
- Flight Number
- Origin Airport
- Destination Airport
- Scheduled & Actual Times
- Arrival Delay
- Departure Delay
- Weather Delay

- Security Delay
- Airline Delay
- Late Aircraft Delay
- Air System Delay
- Cancelled Indicator
- Cancellation Reason
- Distance

Dimension Table:

- Date
- Year
- Quarter
- Month
- Month Name
- Airline Code
- Airline Name
- Cancellation Code
- Cancellation Description
- IATA_CODE
- Airport Name
- City
- State
- Country
- Latitude
- Longitude