Yardgreen - Data Science Integrated Internship

Week 2: project link: Click here

Data Cleaning and Preparation

Objective: Ensure the dataset is clean and ready for analysis

1. **Missing Values Report**:

Rows: 48,120

Columns: 4

Columns:

- DateTime → Timestamps (currently as object/string)
- Junction → ID of traffic junction
- Vehicles → Number of vehicles
- ID → Record ID

Cleaning Plan

- No missing values found in any column.

2. Standardized Date and Time Format:

Standardize Date/Time

Convert DateTime from string to proper datetime format YYYY-MM-DD HH:MM:SS

Add derived columns:

- Date
- Hour
- DayOfWeek
- Month

Clean and consistent `DateTime`
Added columns for flexible time-based analysis
Visual-ready structure for grouping, filtering, and trend analysis

3. Outlier Detection and Removal Summary:

Eliminate Outliers (Using IQR Method)

$IQR = Q3 - Q1$

$\text{Lower Bound} = Q1 - 1.5 \times IQR$

$\text{Upper Bound} = Q3 + 1.5 \times IQR$

Upper Bound (Q3 + 1.5*IQR): 59 vehicles

Remove the outliere in vihicles

4. Refined, Clean Dataset:

Accurate Time-Based Analysis

- With standardized DateTime, you can now group data by hour, day, week, or month to see traffic trends clearly.
- Enables robust visuals like line charts, area charts, and heatmaps.

Cleaner Statistical Summaries

- Outlier removal ensures means, medians, and totals reflect actual traffic behavior.
- Prevents skewed insights from rare traffic spikes or data errors.

Reliable Filtering and Forecasting

- No missing values = no NULL errors in DAX calculations or visuals.
- Clean structure allows smoother integration into forecasting models or Power BI Q&A.

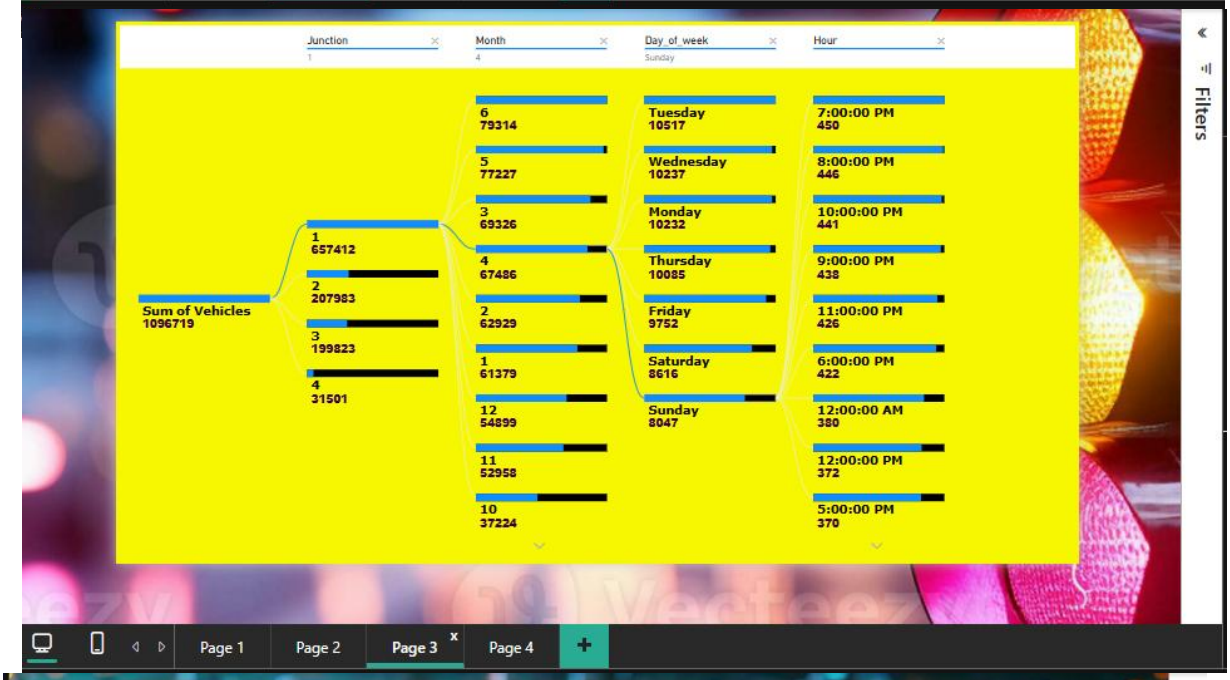
Ready for Dashboards & KPIs

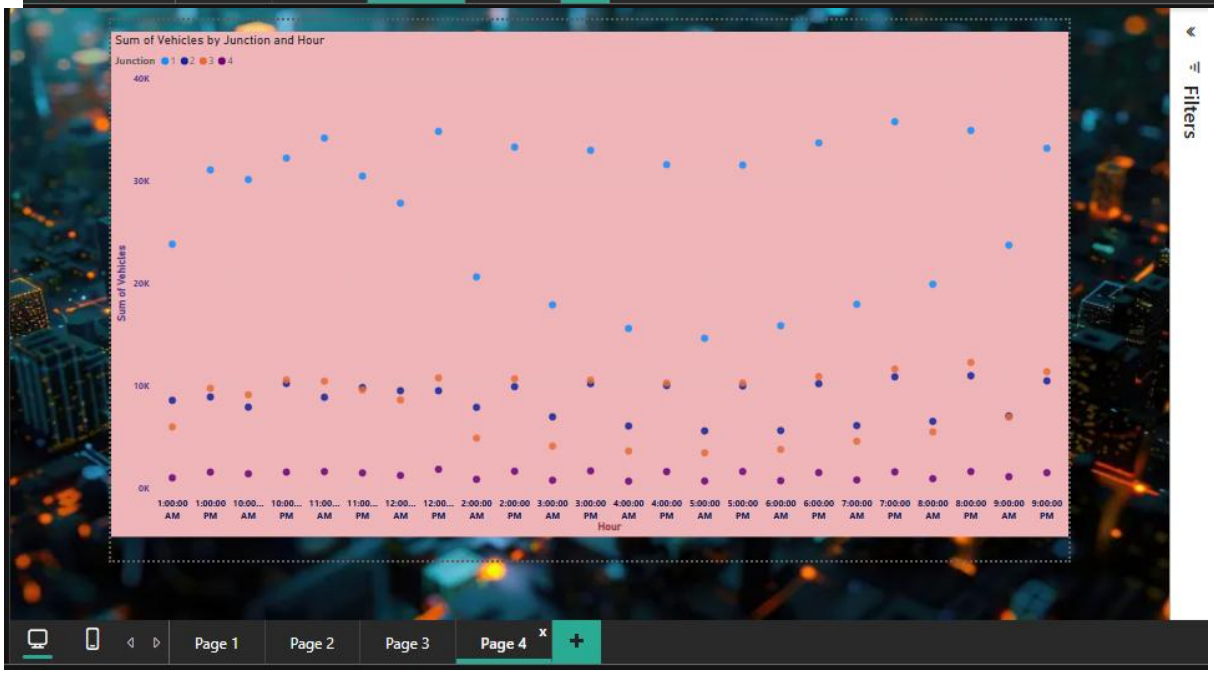
- Supports creation of accurate KPIs, cards, trend lines, and alerts

Visualize


Yardgreen - Data Science Integrated Internship
Week 2: Data Cleaning and Preparation
Filters
Page 1
Page 2
Page 3
Page 4


Cross vehicle
48.12K
No of Junction
4
Sum of cross Vehicles by Year, Quarter, Month and Day
Sum of cross Vehicles by Hour and Day_of_week
Sum of Vehicles by Junction and Junction
Sum of Vehicles by Day_of_week
Filters
Page 1
Page 2
Page 3
Page 4


Junction
Month
Day_of_week
Hour
Sum of Vehicles
1096719
1
657412
2
207983
3
199823
4
31501
6
79314
5
77227
3
69326
4
67486
2
62929
1
61379
12
54899
11
52958
10
37224
Tuesday
10517
Wednesday
10237
Monday
10232
Thursday
10085
Friday
9752
Saturday
8616
Sunday
8047
7:00:00 PM
450
8:00:00 PM
446
10:00:00 PM
441
9:00:00 PM
438
11:00:00 PM
426
6:00:00 PM
422
12:00:00 AM
380
12:00:00 PM
372
5:00:00 PM
370
Filters
Page 1
Page 2
Page 3
Page 4


Sum of Vehicles by Junction and Hour
Junction
Sum of Vehicles
Hour
Filters
Page 1
Page 2
Page 3
Page 4