London Road Collision Analysis (2014-2023)

A Data-Driven Approach to Improving Road Safety

The Challenge: Enhancing Road Safety in London

Scenario: The task at hand was to analyze a decade of road collision data for London.

Objective: To transform raw government data into a clean, reliable data warehouse. Identify high-risk areas, understand patterns, and provide actionable insights.

Justification: Data-driven decision-making is critical for effective resource allocation and policy-making in a complex city like London.

Guiding the Analysis: Key Business Questions

What is the total collision count by month and London Borough (2014-2023)?

What are the top 5 vehicle types involved in fatal collisions?

How does collision severity compare between day and night in non-built-up areas?

What is the risk of serious/fatal collisions during bad weather in Inner vs. Outer London?

What is the year-over-year trend in cyclist and pedestrian casualties during rush hour?

Technical Architecture & Methodology

Methodology:	• Adopted key principles from Ralph Kimball's data warehouse lifecycle to ensure business alignment.
Data Sources:	• DfT Road Safety Data (Primary), Ordnance Survey (Geospatial), ONS (Statistical).
Technology Stack:	• R Programming for ETL • Hadoop/HDFS & Hive for scalable storage and query. • Tableau for Visualisations

Data Engineering

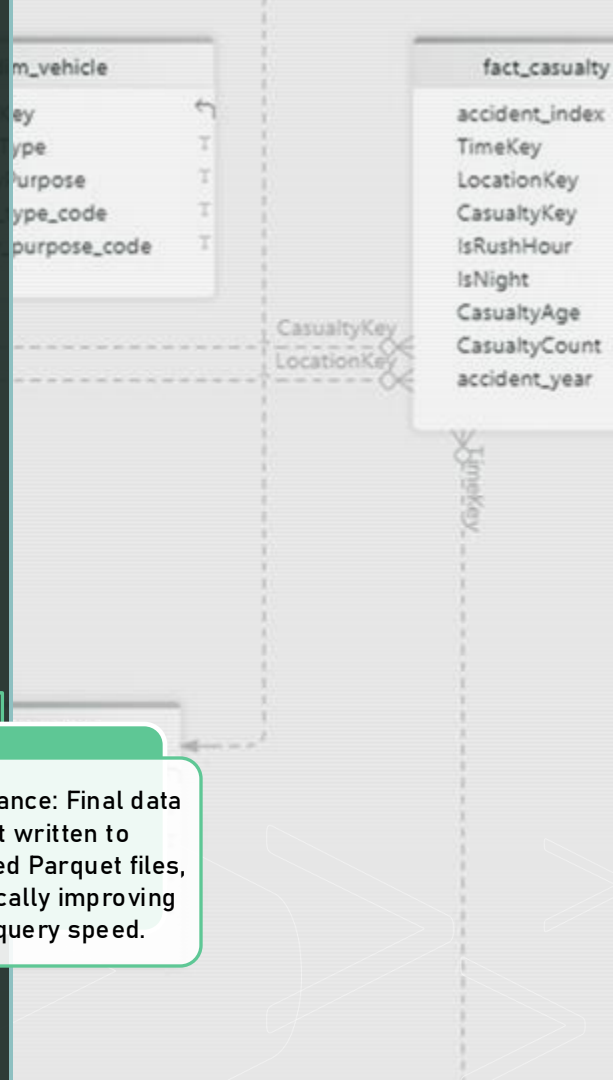
Data Modeling: Designed a Fact Constellation Schema

ETL Process Highlights:

Safe Parsing & Cleaning: Filtered over 1,300 implausible age records and standardized missing values to true NULLs.

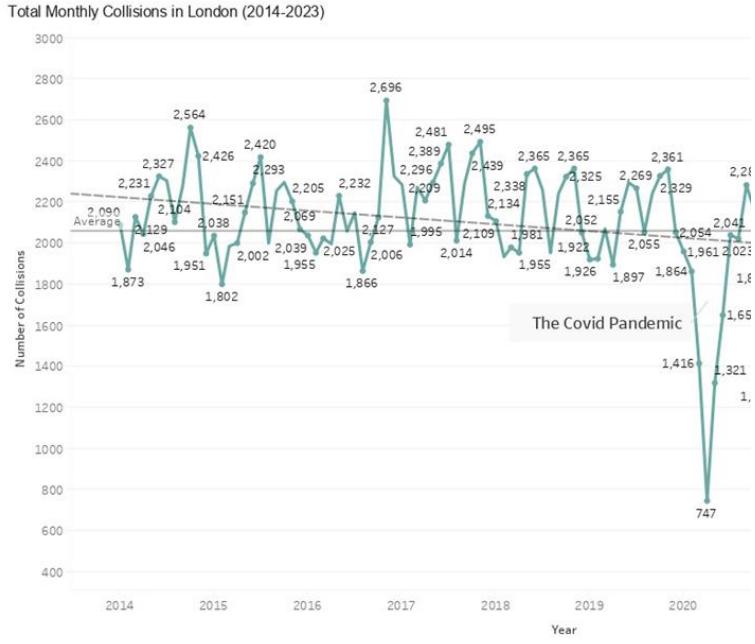
Enrichment: Performed spatial join (IsBuiltUp flag) and statistical joins (Population/Area for normalization).

Performance: Final data mart written to partitioned Parquet files, dramatically improving Hive query speed.

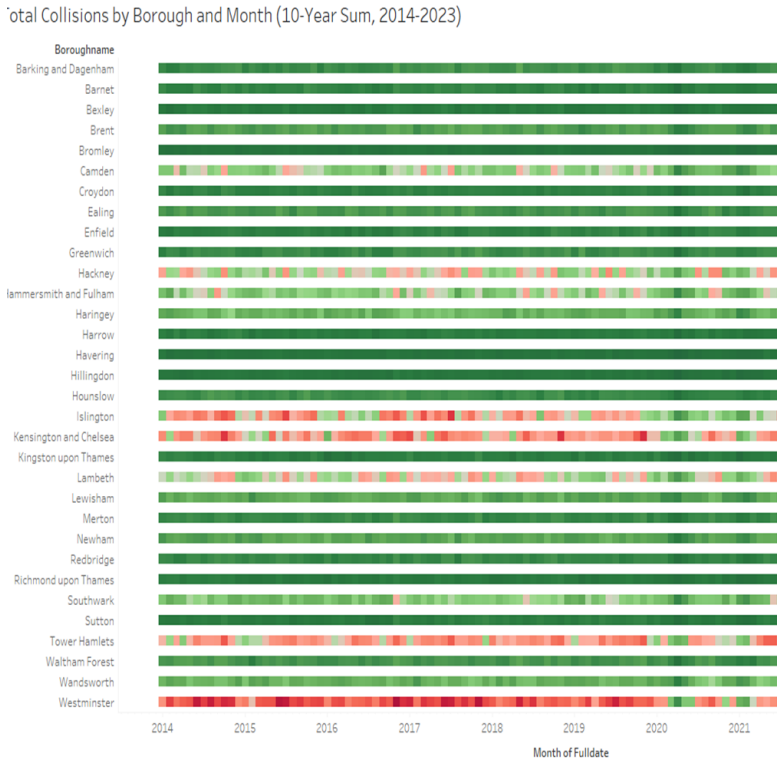


Temporal & Normalized Geographic Hotspots

- **Temporal Trends: Collisions consistently peak in autumn and on Fridays.**
- **Geographic Hotspots: Normalizing by land area reveals the true high-risk areas are central, high-density boroughs (City of London, Islington, Kensington & Chelsea), not just the largest boroughs by volume.**



The trend of sum of Collisioncount for Fulldate Month. The marks are labelled by sum of Collisioncount. The data is filtered on Fullda



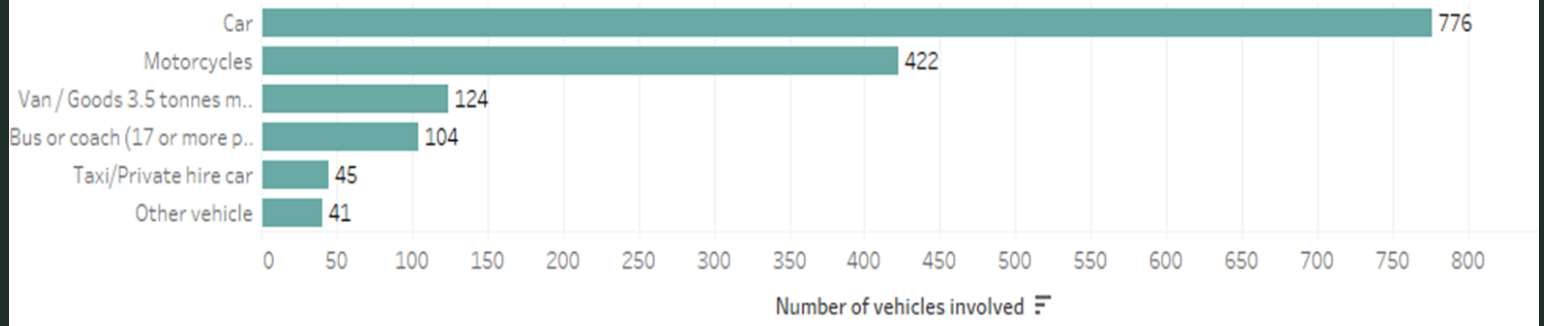
illdate Month for each Boroughname. Colour shows collision count/sqkm (sandbox-hdp.hortonworks.com). The view is filtered on Boroughname, which excludes City of London.

Fatal Risk: Cars are overwhelmingly involved in fatal incidents (776 over 10 years).

Prioritization: Motorcycles represent a disproportionate risk, involved in 422 fatal collisions. They must be a primary focus for safety interventions.

Fatal Collision Vehicle Analysis

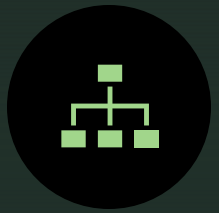
Top Vehicle Types Involved in Fatal Collisions (2014-2023)



Sum of Vehicleinvolvementcount for each Motorcycles. The marks are labelled by sum of Vehicleinvolvementcount. The data is filtered on Casualtyseverity and Vehicletype. The Casualtyseverity filter keeps Fatal. The Vehicletype filter keeps 10 of 22 members.

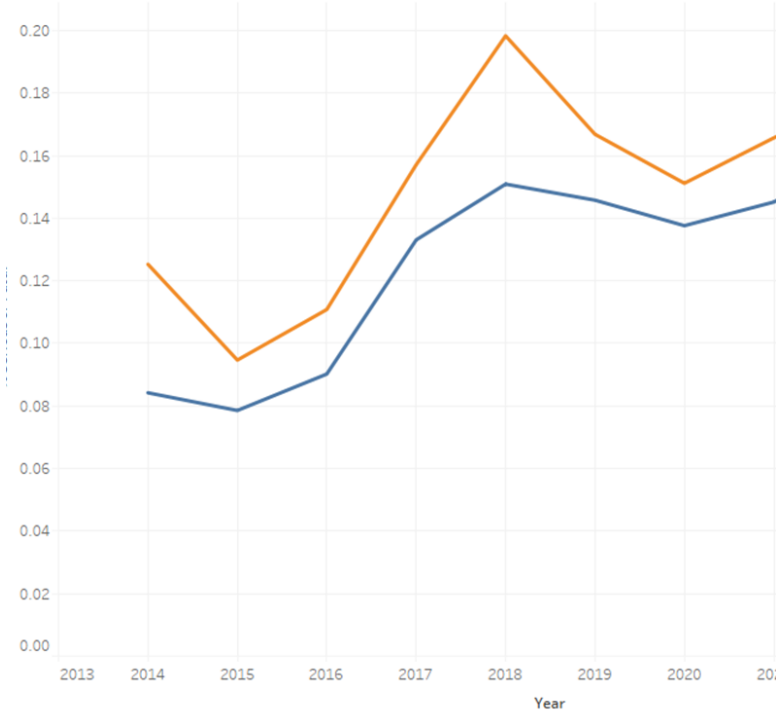
Vulnerable Users & Environmental Factors

Vulnerable Users: Casualties for cyclists and pedestrians are returning to pre-pandemic volumes.

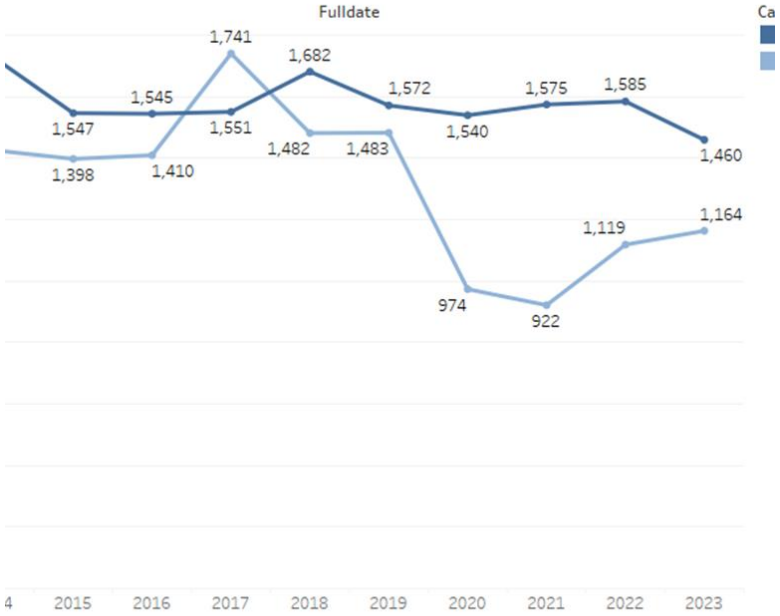


Night-Time Risk Multiplier: In less-developed, non-built-up areas, the risk of a collision being serious or fatal is consistently higher at night, confirming poor visibility is a major severity factor.

ay vs Night



user trends

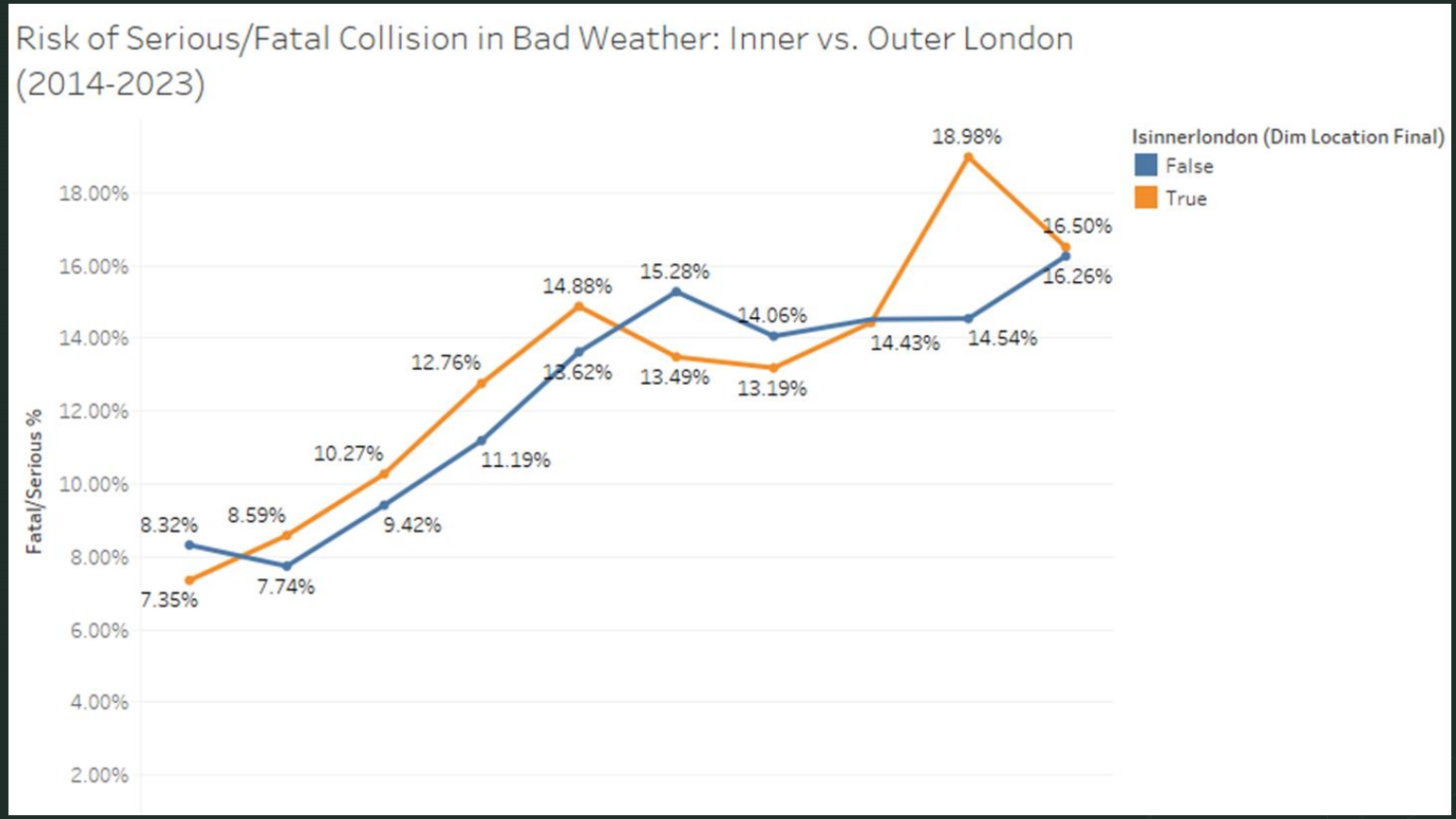


f Casualtycount (Fact Casualty) for Fulldate Year. Colour shows details about Casualtyclass. The
by sum of Casualtycount (Fact Casualty). The data is filtered on Isbuiltup (Dim Location Bua),
asualty) and Casualtytype. The Isbuiltup (Dim Location Bua) filter keeps True. The Isrushhour (Fact
ps True. The Casualtytype filter keeps Cyclist and Pedestrian. The view is filtered on Casualtyclass,
or rider and Pedestrian.

Adverse Weather & Severity Risk

DENSITY MULTIPLIER: DURING ADVERSE WEATHER (RAIN/SNOW), THE RISK OF SEVERE/FATAL COLLISIONS IS CONSISTENTLY HIGHER IN INNER LONDON COMPARED TO OUTER LONDON.

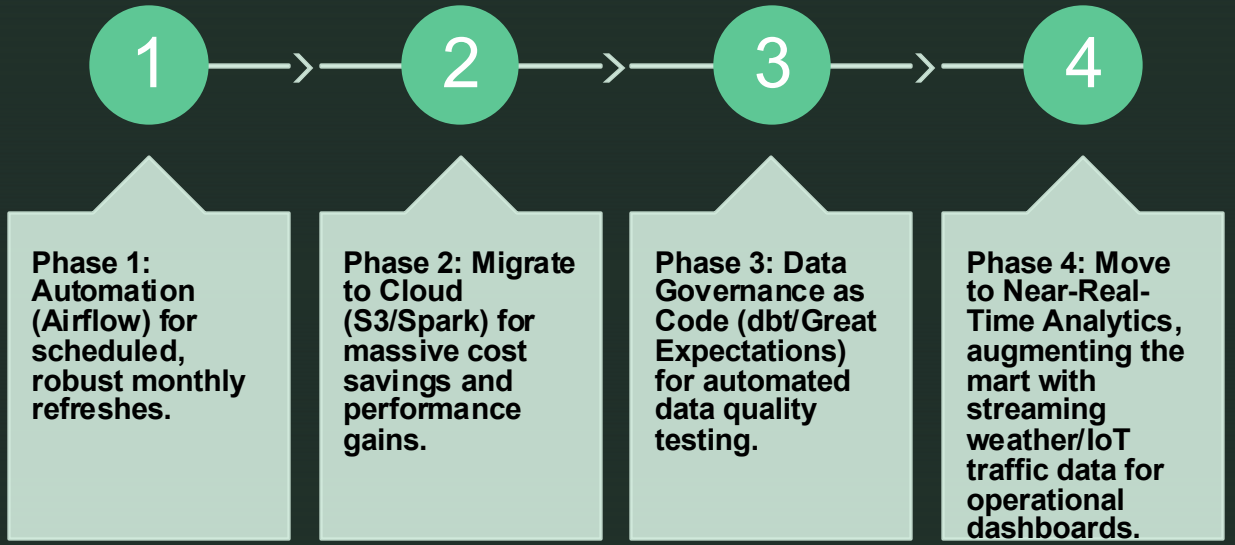
THIS PEAKED NEAR 19% IN 2022 FOR INNER LONDON, SUGGESTING THAT THE COMBINATION OF POOR WEATHER AND HIGHER CENTRAL TRAFFIC DENSITY LEADS TO COMPOUNDING SEVERITY.


Risk of Serious/Fatal Collision in Bad Weather: Inner vs. Outer London (2014-2023)
Isinnerlondon (Dim Location Final)
False
True
Fatal/Serious %
18.00%
16.00%
14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
8.32%
7.74%
9.42%
11.19%
13.62%
15.28%
14.06%
14.43%
14.54%
16.26%
7.35%
8.59%
10.27%
12.76%
14.88%
13.49%
13.19%
18.98%
16.50%

Actionable Insights & Recommendations

Target	Target High-Density Corridors: Focus enforcement in City of London/Islington during high-risk autumn peaks.
Prioritize	Prioritize Motorcycle Safety: Dual-focus campaign on rider training and car driver awareness.
Enhance	Enhance Night-Time Infrastructure: Conduct formal review of street lighting and road markings in outer boroughs.
Implement	Implement Weather-Responsive Ops: Use VMS alerts in Inner London during rain/snow to warn drivers of heightened severity risk.

Future Development & Operationalization Roadmap



Any questions, feel free to reach out at
syedtazeem40@gmail.com