Smart City Data Integration Project - Summary

# Project Objective

The objective of this project is to integrate various data sources (traffic, public services, pollution) to optimize city planning and enhance urban living.

# Methodology

The project utilized several advanced data science techniques including K-Means clustering to segment days based on city dynamics, Isolation Forest for anomaly detection in pollution levels, and regression analysis to explore the relationship between traffic and public service usage. The analyses were complemented by geographic heatmaps and time series analysis to identify trends and patterns.

# Data Science & Analytics

## 1. Clustering Analysis

Days were segmented into clusters based on traffic congestion, public service usage, and pollution levels using K-Means clustering. This helped identify distinct patterns in city dynamics.

## 2. Anomaly Detection

An Isolation Forest model was used to detect anomalies in pollution levels. This analysis identified outliers that could indicate potential environmental issues.

## 3. Regression Analysis

A regression analysis explored the relationship between traffic congestion and public service usage. The model revealed significant correlations, suggesting that increased traffic congestion leads to higher public service usage.

## 4. Geographic Heatmap

A geographic heatmap was created to visualize pollution levels across different locations. This visualization helped identify areas with higher pollution levels, which can inform targeted interventions.

## 5. Time Series Analysis

Time series analysis was conducted to explore trends in traffic congestion and public service usage over time. This analysis highlighted periods of high congestion and service demand, which can inform resource allocation.

# Summary and Recommendations

The analysis provided valuable insights into the factors influencing city dynamics. Clustering and anomaly detection identified patterns and outliers, while regression analysis revealed significant relationships between variables. These insights can be used to optimize city planning and enhance urban living.

We recommend implementing real-time monitoring and clustering of traffic and pollution data to optimize urban planning. Anomaly detection should be used to quickly identify and address environmental and public service issues. Regression analysis can be leveraged to predict the impact of traffic on public services and adjust resource allocation accordingly.