

## **Analyzing public transportation:**



**Ridership Patterns:** Determine peak and off-peak ridership times and locations to optimize service frequency and resource allocation.

**Route Efficiency:** Evaluate the efficiency of existing routes to identify areas for route optimization or expansion based on passenger demand.

**Customer Satisfaction:** Measure customer satisfaction through surveys and feedback to identify areas for improvement in service quality and amenities.

### **On-Time Performance:**

**Punctuality Metrics:** Assess the percentage of trips that arrive on time, with specific benchmarks for peak and off-peak hours.

**Delay Causes:** Identify common causes of delays (e.g., traffic, mechanical issues) to prioritize improvements.

**Schedule Adherence:** Measure how closely actual departure and arrival times align with the published schedule.

### **Passenger Satisfaction:**

Conduct regular passenger satisfaction surveys to gauge their experiences and identify pain points.

**Amenities Evaluation:** Assess passenger satisfaction with amenities like cleanliness, accessibility, and information displays.

**Complaint Analysis:** Analyze and categorize passenger complaints to address recurring issues and improve service quality.

### **Service Efficiency:**

**Load Factor:** Monitor the capacity utilization of vehicles to optimize scheduling and resource allocation.

**Service Frequency:** Evaluate whether service frequency aligns with demand, avoiding overcrowding or underutilization.

**Travel Time:** Measure the average travel time for passengers and identify areas for route optimization or traffic management.

### **Operational Cost Reduction:**

**Fuel Efficiency:** Analyze fuel consumption data to identify opportunities for cost-saving measures.

**Maintenance Efficiency:**

Assess maintenance schedules and costs to optimize vehicle upkeep without compromising safety.

### **DATA COLLECTION:**

Transportation data, including schedules, real-time updates, and passenger feedback, can be collected from various sources and using several methods. Here's an overview:

### **Sources for Transportation Data:**

**Transit Agencies:** Public transportation agencies maintain schedules, route information, and operational data, often available on their official websites.

**GPS and Tracking Systems:** These systems are installed on vehicles to provide real-time location updates, enabling real-time tracking of buses, trains, or other transit modes.

**Mobile Apps:** Many transit agencies provide mobile apps that offer real-time information on schedules, delays, and service alerts. Apps can also collect data from users, such as their location and trip preferences.

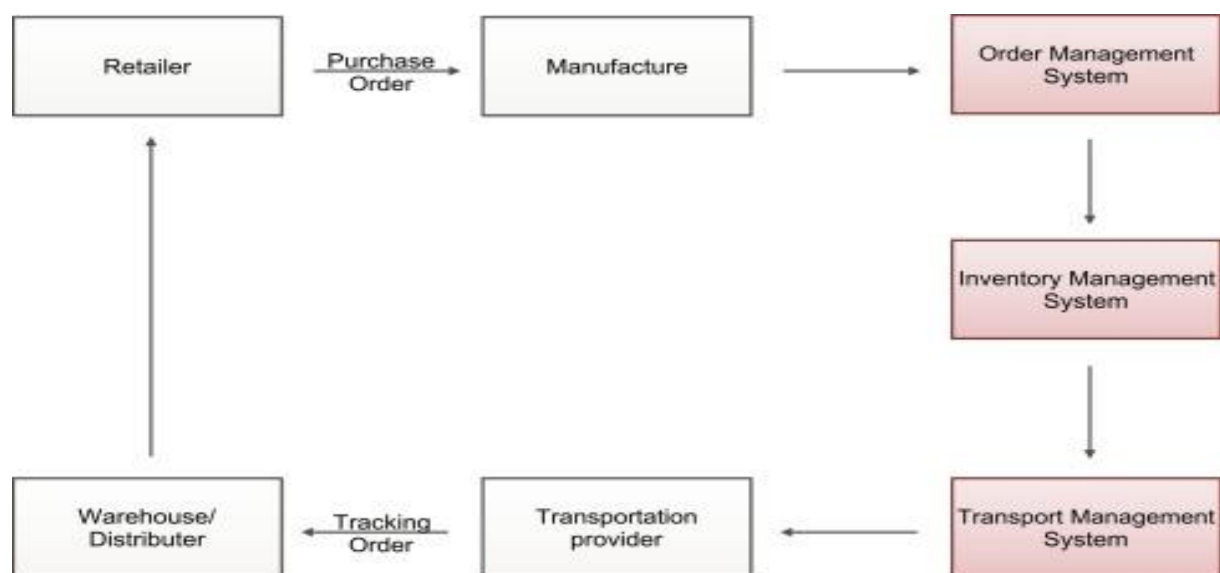
**Automated Fare Collection:** Systems like smart cards and contactless payments capture passenger boarding and alighting data, helping to track ridership and travel patterns.

**Traffic Management Systems:** Data from traffic management centers, such as traffic cameras and road sensors, can impact transit schedules and real-time updates.

**Passenger Counting Systems:** These systems, often using sensors at entry and exit points, count passengers to assess load factors and demand patterns.

**Surveys and Feedback Forms:** Conducting surveys, both in-person and online, and providing feedback forms to passengers can gather valuable insights into their experiences and preferences.

**IoT Devices:** Internet of Things (IoT) devices can collect data on vehicle conditions, environmental factors, and passenger behavior.



## **Methods for Collecting Transportation Data:**

**Automated Data Collection:** This includes automatic vehicle location (AVL) systems, fare collection systems, and sensor networks that collect data without human intervention

**Mobile Apps and Websites:** Transit agencies can develop mobile apps and websites where passengers can access schedules, receive real-time updates, and provide feedback.

**Smart Cards and Contactless Payments:** Fare collection systems using smart cards or contactless payments can automatically record passenger boarding and alighting, helping to track ridership.

**Surveys and Questionnaires:** Conducting surveys on vehicles, at stations, or online can gather information on passenger satisfaction, travel behavior, and preferences.

**Media and Online Reviews:** Monitoring social media platforms and online review websites can provide real-time feedback and insights into passenger experiences.

**Crowdsourcing:** Engaging the community in collecting data through mobile apps or dedicated platforms can provide real-time updates on transit conditions and incidents.

**Traffic Cameras and Sensors:** Data from traffic cameras and road sensors can be integrated to provide real-time traffic conditions that affect transit schedules.

**Electronic Ticketing Machines:** Data collected through ticketing machines can include passenger counts, fare revenue, and trip data.

**Vehicle and Equipment Sensors:** Sensors on vehicles and infrastructure can monitor conditions, such as engine performance, temperature, and maintenance needs.

**Focus Groups and Interviews:** In-depth interviews and focus group discussions with passengers can provide qualitative insights into their experiences and needs.

## **Visualization Strategy:**

Creating informative dashboards and reports using IBM Cognos to visualize insights from transportation data can be a powerful way to convey information effectively. Here's a plan to achieve this:

### **Define Objectives and Audience:**

Clearly define the objectives of your dashboards and reports. What insights are you trying to convey? Who is your target audience (e.g., transportation officials, management, passengers)?<sup>2</sup>.

### **Data Preparation:**

Ensure your transportation data is cleaned, structured, and stored in a suitable data source that Cognos can access (e.g., a relational database).

### **Connect to Data Sources:**

Use IBM Cognos to connect to your data sources, which can include databases, spreadsheets, and web services.

### **Design Data Models:**

Create data models that define relationships between different data tables, making it easier to pull data for visualization.

### **Dashboard Design:**

Design your dashboard layout. IBM Cognos provides drag-and-drop functionality for arranging visual elements. Consider the placement of key performance indicators (KPIs), charts, graphs, and maps.

### **Select Visualization Tools:**

Choose appropriate visualization tools based on the type of data and insights you want to convey. Options in Cognos include bar charts, line charts, heat maps, and more.

### **Create Interactive Filters:**

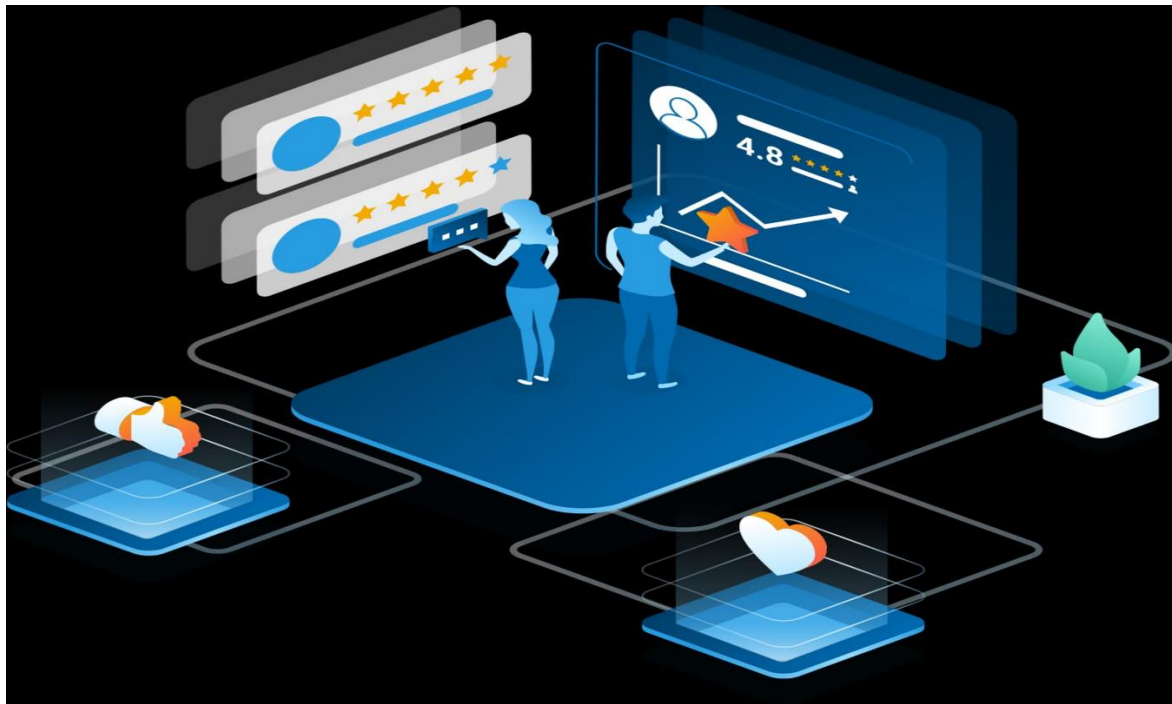
Implement interactive filters that allow users to drill down into specific data points or time frames for deeper analysis.

## **Data Aggregation and Calculation:**

Use Cognos to aggregate and calculate data, such as average travel times, ridership trends, or cost efficiencies.

## **Code Integration:**

Code can enhance various aspects of the analysis of public transportation data, particularly in data preprocessing, transformation, and statistical analysis. Here are the aspects that can benefit from coding:



## **Data Cleaning:**

### **Outlier Detection:**

Code can help identify and handle outliers in ridership data, which might distort statistical analysis.

### **Missing Data Handling:**

Automation can fill in missing data points using imputation techniques or flag them for further investigation.

### **Data Validation:**

Code can validate data for accuracy, consistency, and adherence to predefined formats.

## **Data Transformation:**

### **Feature Engineering:**

Create new features or variables from existing data to capture more relevant information, e.g., calculating daily or hourly averages.

### **Normalization and Scaling:**

Code can normalize data to ensure that variables are on a consistent scale for statistical analysis. One -Hot Encoding: Convert categorical variables into numerical format for modeling purposes.

## **Statistical Analysis:**

### **Descriptive Statistics:**

Automate the calculation of descriptive statistics like mean, median, variance, and percentiles for key performance indicators.

### **Hypothesis Testing:**

Code can facilitate hypothesis testing to determine if there are statistically significant differences in ridership patterns, service efficiency, or passenger satisfaction.

### **Regression Analysis:**

Utilize code to perform regression analysis to understand relationships between variables, such as how schedule changes impact ridership.

### **Time Series Analysis:**

Automate time series analysis to detect trends, seasonality, and forecast future ridership.

### **Spatial Analysis:**

Implement code for spatial analysis, such as identifying optimal locations for new transportation hubs based on geographic data.