

Question 1: What is Anomaly Detection? Explain its types (point, contextual, and collective anomalies) with examples.

Answer: Anomaly Detection identifies unusual patterns or points in data that deviate from expected behavior.
Types: Point anomaly (single outlier), Contextual anomaly (context-dependent, e.g., high temp in winter), Collective anomaly (a sequence anomalous together, e.g., zero transactions in an active account).

Question 2: Compare Isolation Forest, DBSCAN, and Local Outlier Factor.

Answer: Isolation Forest isolates anomalies using random partitions (good for high dimensions). DBSCAN detects anomalies as points in low-density regions (good for spatial/arbitrary shapes). LOF measures density deviation compared to neighbors (good for local anomalies).

Question 3: Key components of Time Series.

Answer: Trend (long-term growth/decline), Seasonality (repeated periodic effects), Cyclic patterns (long-term fluctuations not fixed), Residual (random noise). Example: AirPassengers shows trend and seasonality.

Question 4: Stationarity in time series.

Answer: A stationary series has constant mean/variance/autocorrelation over time. Tests: ADF, KPSS.
Transformations: differencing, log transform, seasonal decomposition, detrending.

Question 5: AR, MA, ARIMA, SARIMA, SARIMAX differences.

Answer: AR uses past values, MA uses past forecast errors, ARIMA combines AR+I(differencing)+MA, SARIMA adds seasonality, SARIMAX adds exogenous predictors.

Question 6: Decompose AirPassengers into trend, seasonality, residuals.

Answer: Using seasonal decomposition, the dataset is split into trend, seasonality, and residual components.

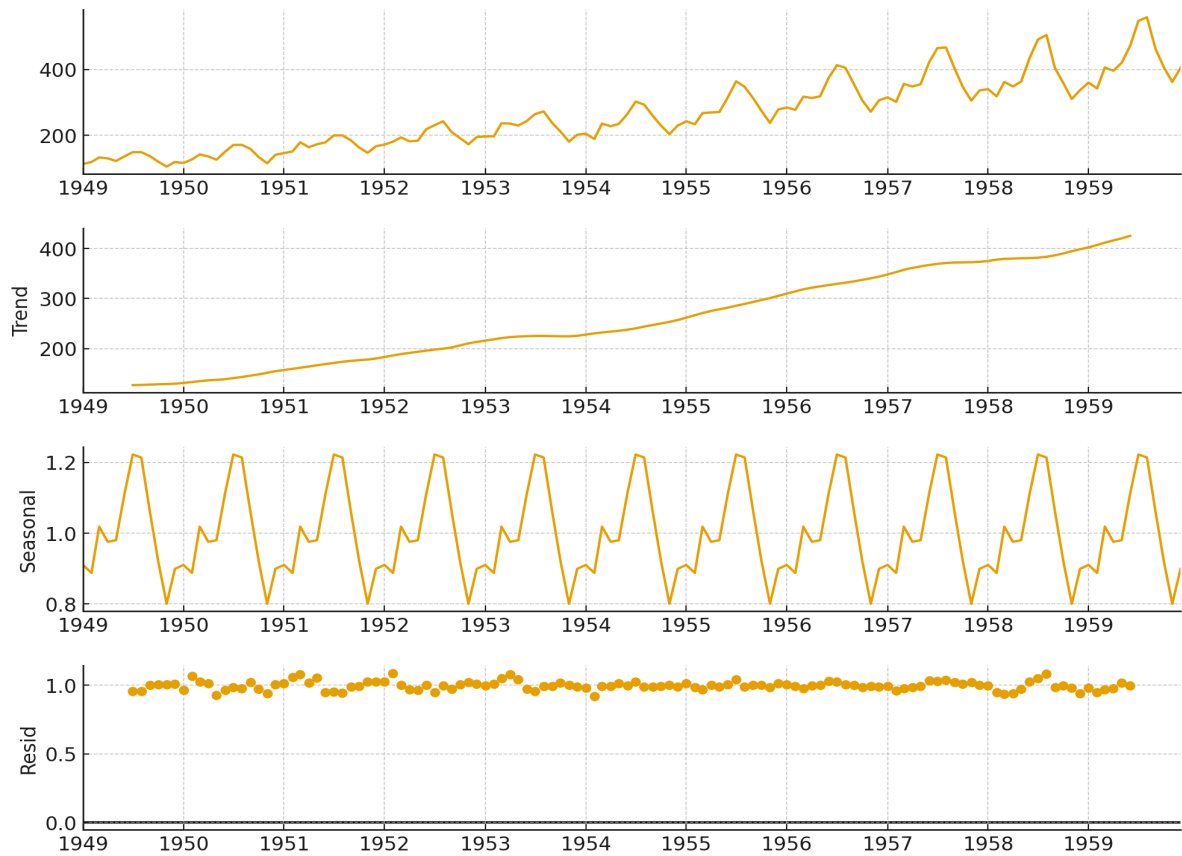


Figure 1: Decomposition of AirPassengers series into observed data, trend, seasonality, and residuals.

Question 7: Isolation Forest on NYC Taxi Fare data.

Answer: Isolation Forest applied to fare amount and passenger count detects anomalies.

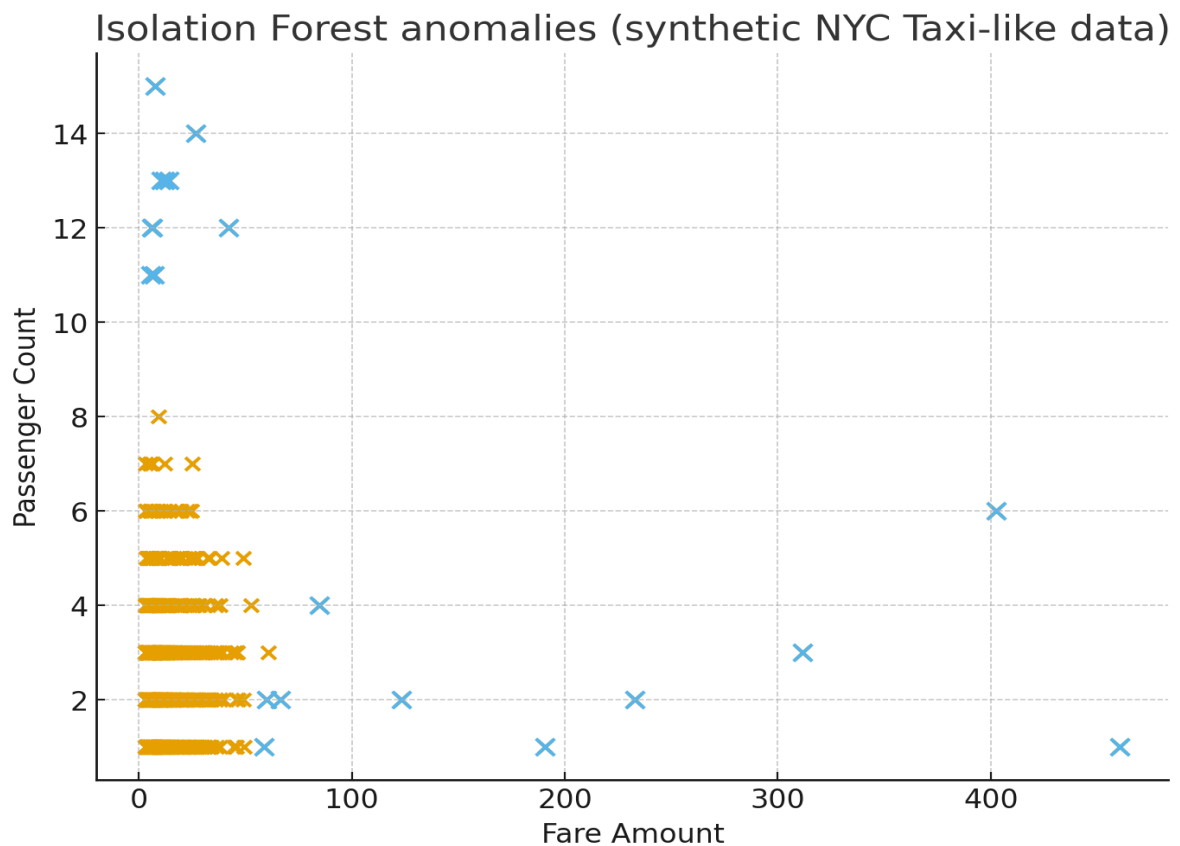


Figure 2: Isolation Forest anomalies highlighted on synthetic NYC Taxi-like dataset.

Question 8: SARIMA forecast on AirPassengers dataset.

Answer: SARIMA(1,1,1)(1,1,1,12) model trained to forecast 12 future months of passenger data.

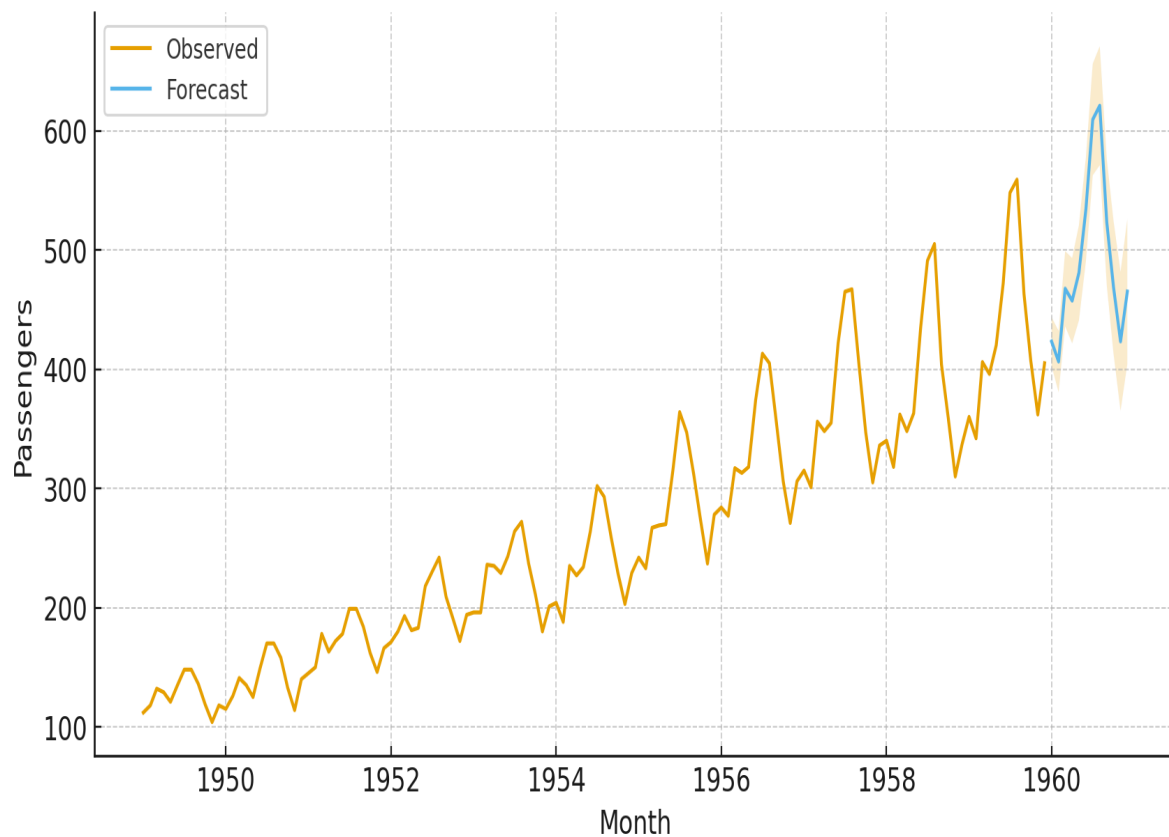


Figure 3: SARIMA model 12-month forecast for AirPassengers series with 95% confidence intervals.

Question 9: Local Outlier Factor (LOF) on Taxi data.

Answer: LOF detects anomalies based on density deviation relative to neighbors.

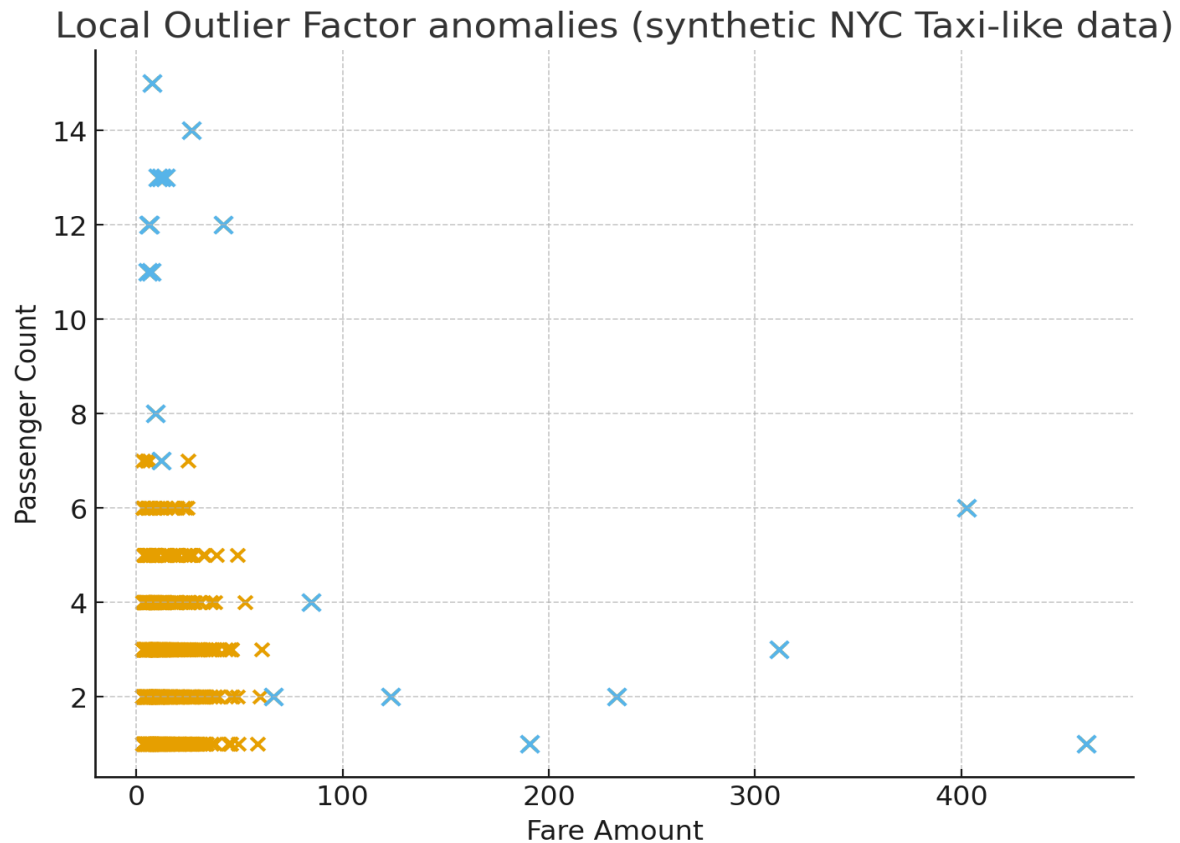


Figure 4: Local Outlier Factor anomalies detected on synthetic NYC Taxi-like dataset.

Question 10: Real-time data science workflow for power grid monitoring.

Answer:

1. Data ingestion: stream data in real-time (Kafka, Spark Streaming).
2. Anomaly detection: use Isolation Forest, LOF, or DBSCAN to catch spikes/drops.
3. Forecasting: apply SARIMAX to capture seasonality and weather effects.
4. Validation & monitoring: rolling forecasts, metrics (RMSE/MAE), drift detection.
5. Business impact: optimize load balancing, prevent outages, improve planning and pricing.

Note: Demonstrations use AirPassengers (classic) and a synthetic NYC Taxi-like dataset. Replace with your actual CSVs for precise outputs.