Anomalies Detection

Introduction

What is Anomaly Detection

- Anomaly detection involves identifying the differences, deviations and exceptions from the norm in a dataset.
- Also called outliers' detection
- Anomaly detection makes this data not only useful but powerful for Algorithm Building

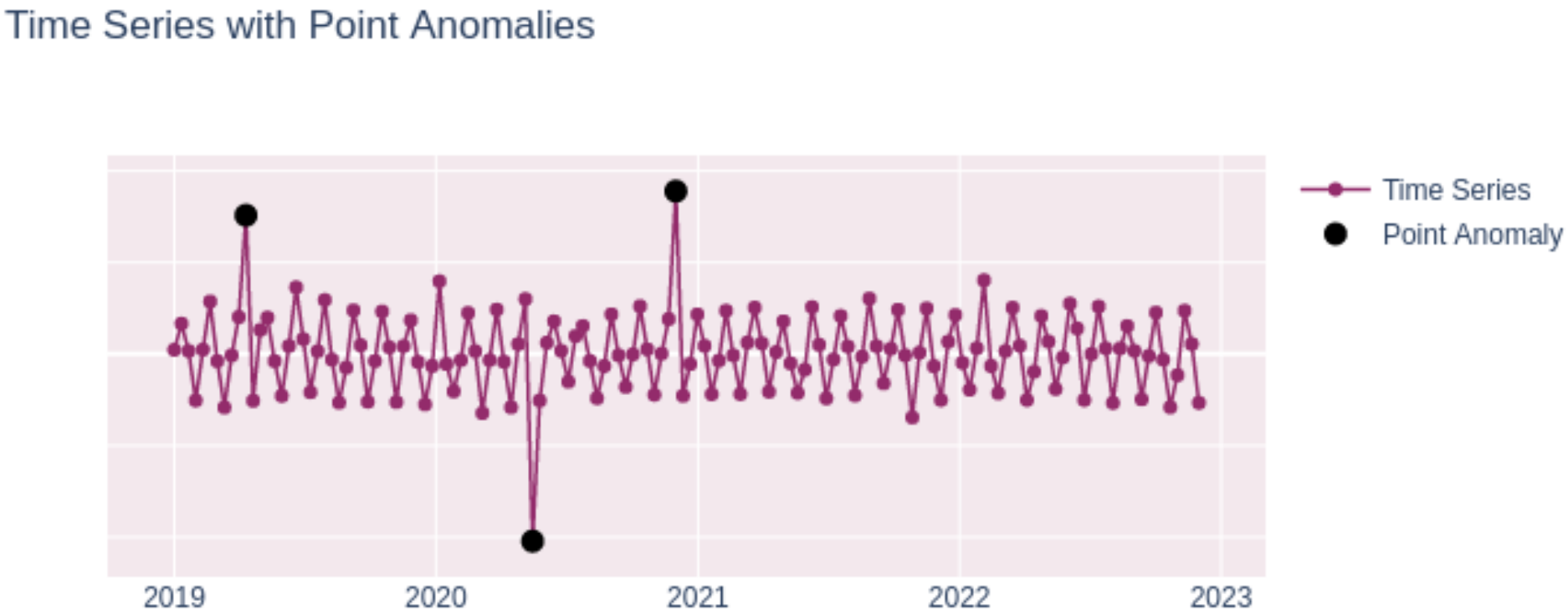
Types in Anomalies Detection

- Point
- Collective
- Contextual

Point Anomalies

- A point anomaly is where a single datapoint stands out from the expected pattern, range, or norm.
- In other words, the datapoint is unexpected.
- Examples can be unusually high values or low values

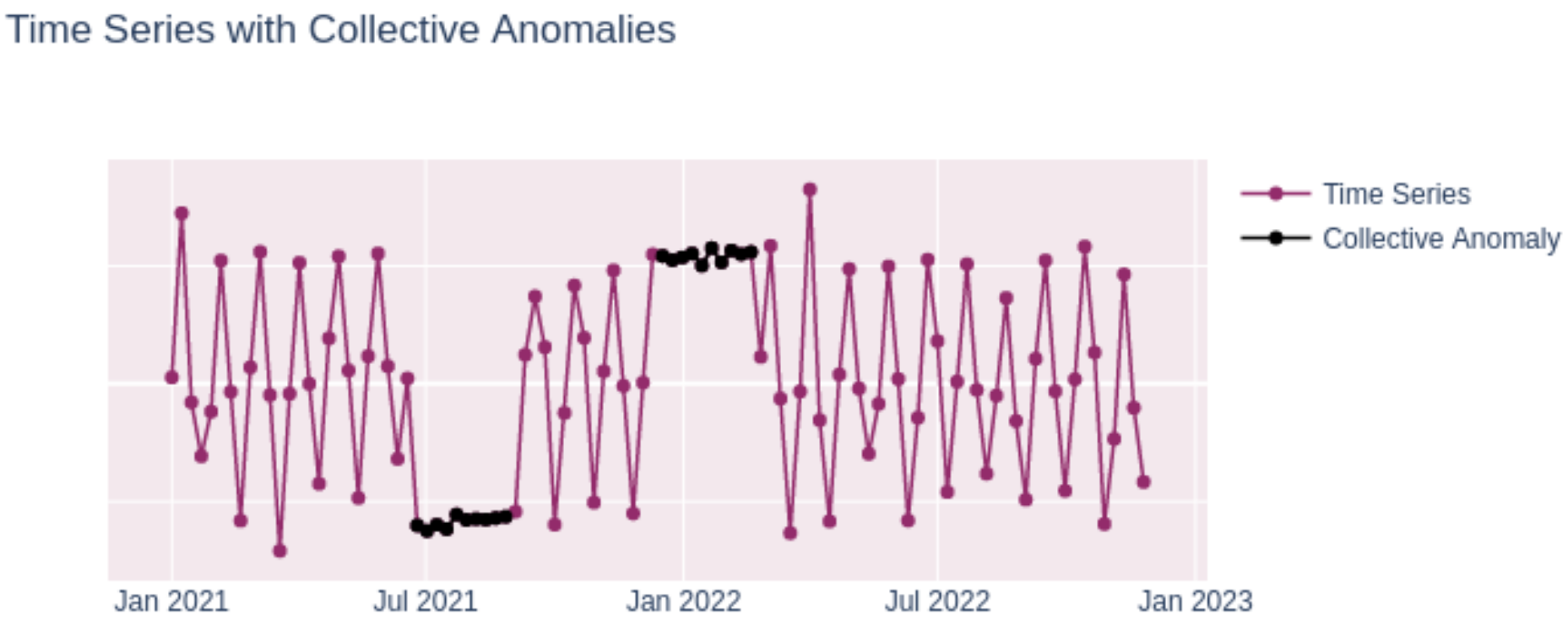
Example of Point Anomalies



Collective Anomalies

- Collective anomalies are defined as sequences of observations that are not necessarily anomalous when considered individually, but together form an anomalous pattern.

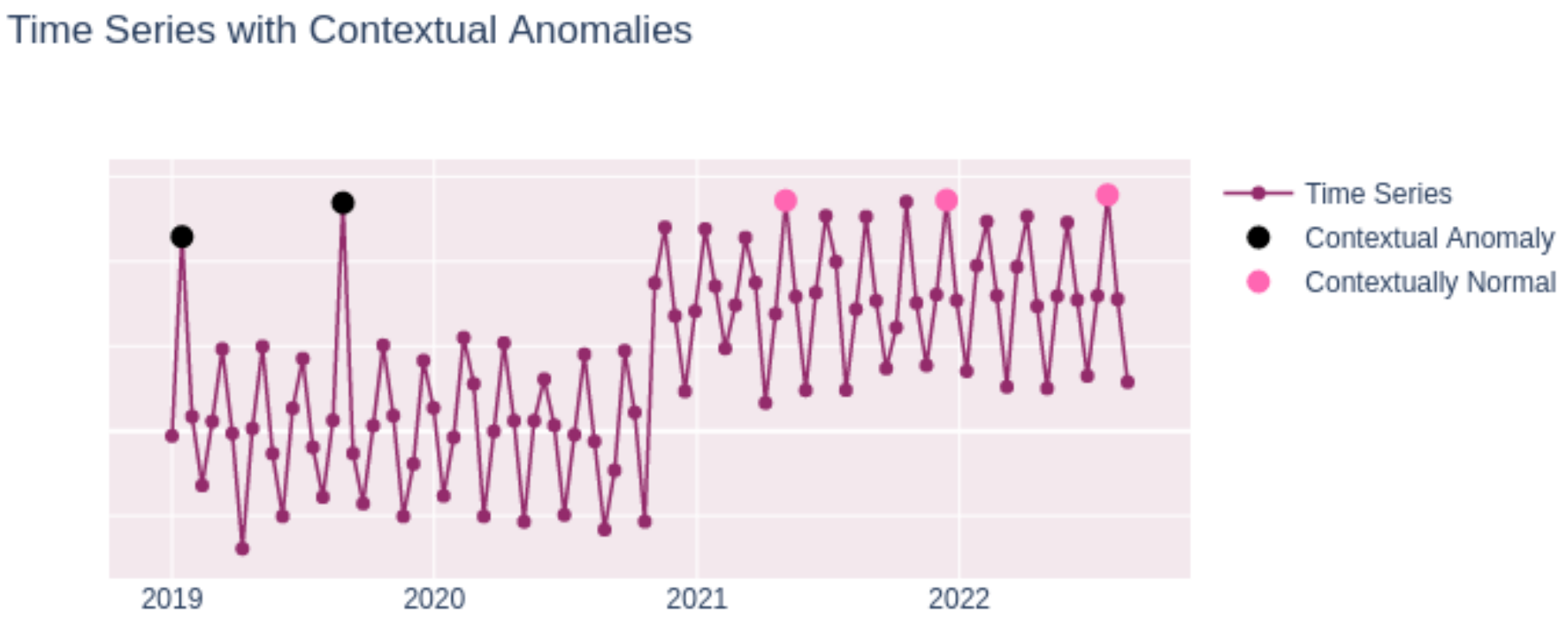
Example of Collective Anomalies



Contextual Anomalies

- Contextual anomalies refer to data points that deviate significantly within a specific context but appear normal outside of that context.
- e.g. High spending during festive occasions may be a routine. But the same high spendings when others have low spendings

Example of Contextual Anomaly



Methods of Detection

- Supervised
- Unsupervised

Supervised Anomalies Detection

- Supervised anomaly detection uses labeled data to train a classifier that can distinguish between normal and anomalous instances.
- The labels indicate whether an instance belongs to the normal class or one of the predefined anomaly classes.

Unsupervised Anomalies Detection

- Unsupervised anomaly detection does not require labeled data to identify outliers.
- Instead, it relies on statistical or distance-based measures to assess how different an instance is from the rest of the data.
- For example, we can use clustering techniques, Isolation Forest