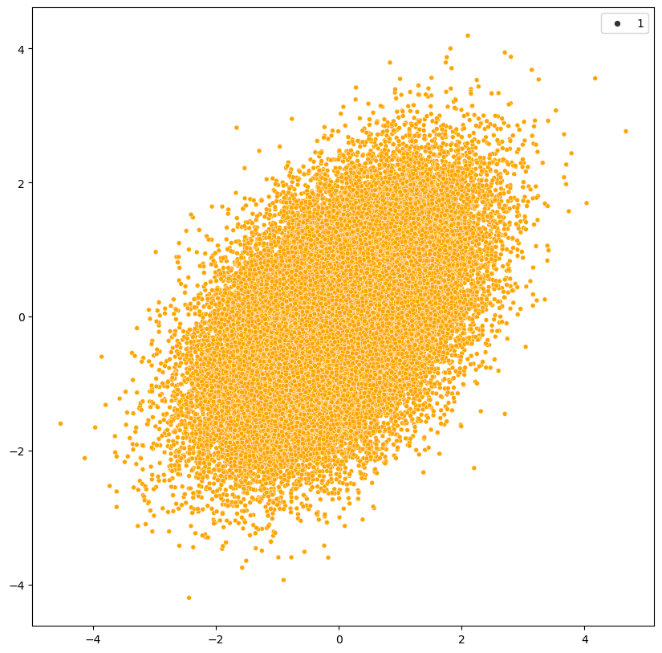
**Outlier Analysis for Normal Data using Scatter Plot**

Note : All the data is having 2 dimension and 50000 data points. Only difference between different data is the value of mean and covariance of attributes

* **Data 1**

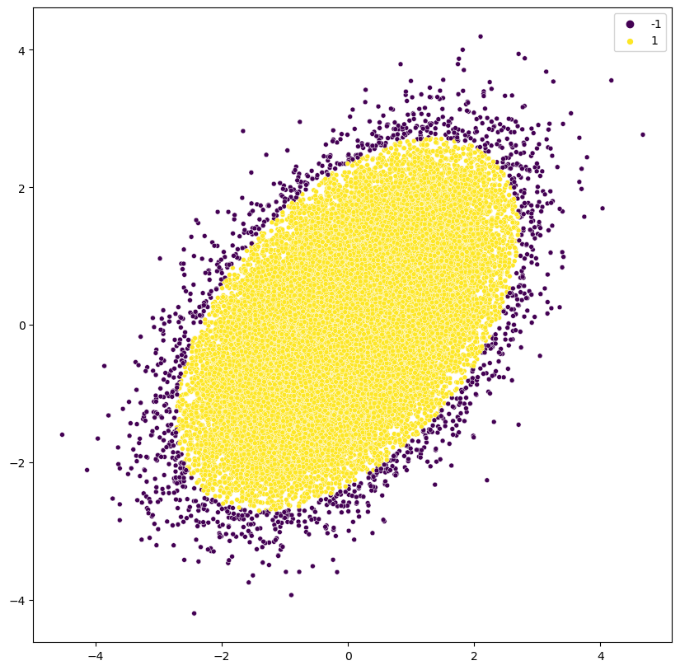
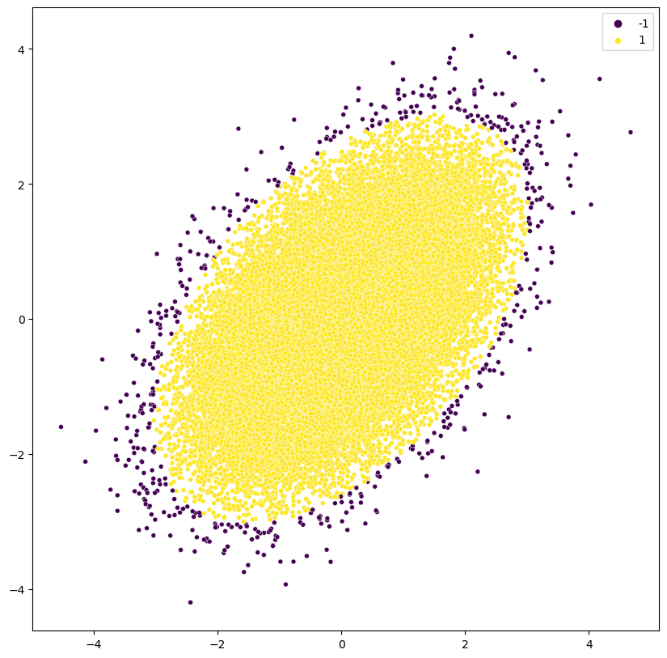
Mean : [0,0]

Covariance : [[1,0.5], [0.5,1]]



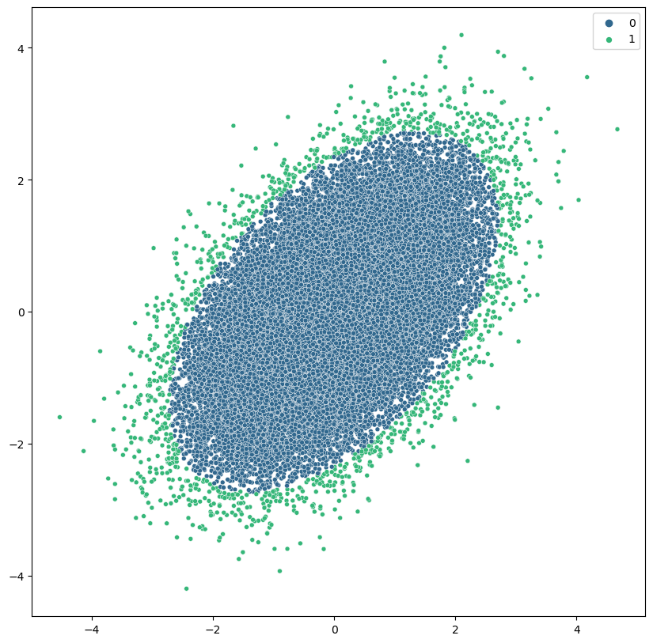
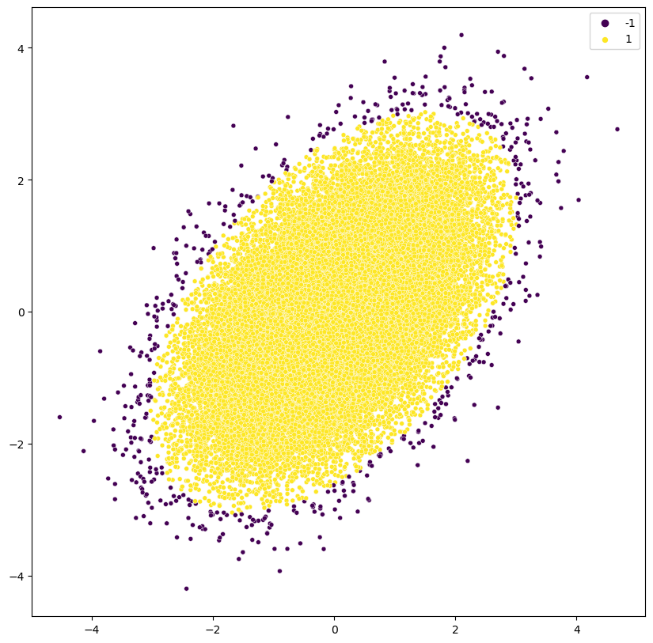
Original Data Scatter Plot

1. Mahalanobis Distance

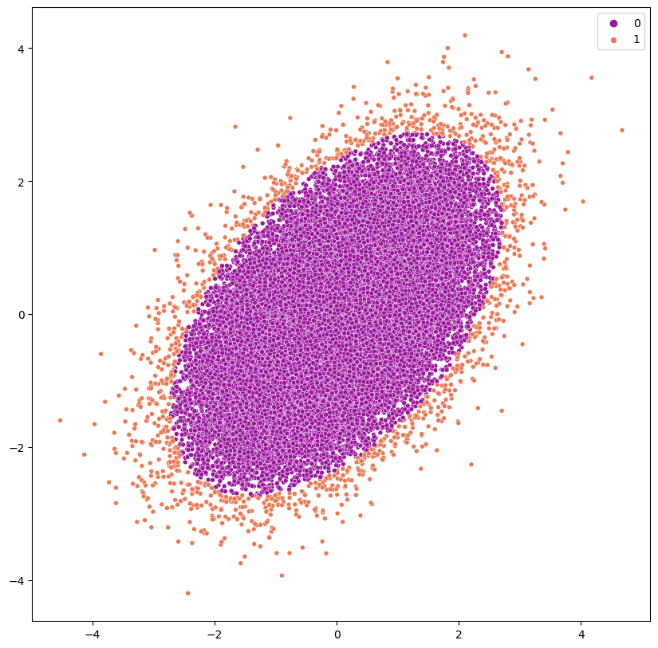
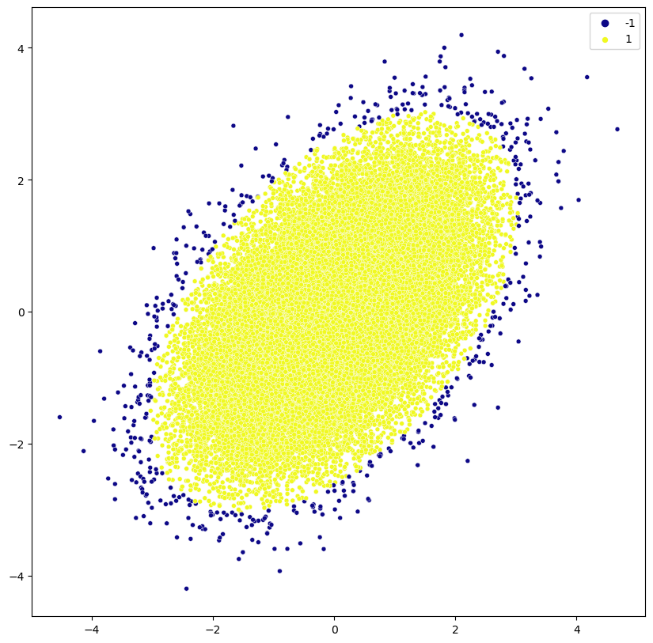
Contamination = 0.025 Contamination = 0.01

1. Elliptic Envelope

Contamination = 0.025 Contamination = 0.01

1. Minimum Covariance Determinant

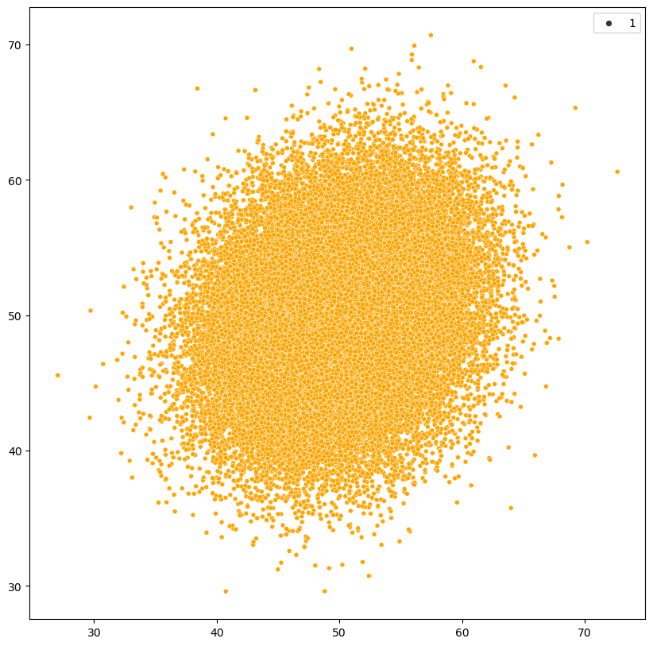
 

Contamination = 0.025 Contamination = 0.01

* **Data 2**

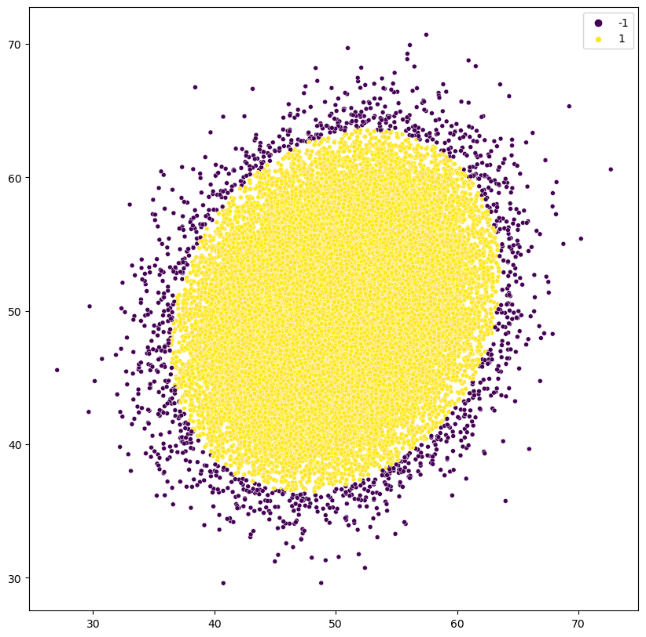
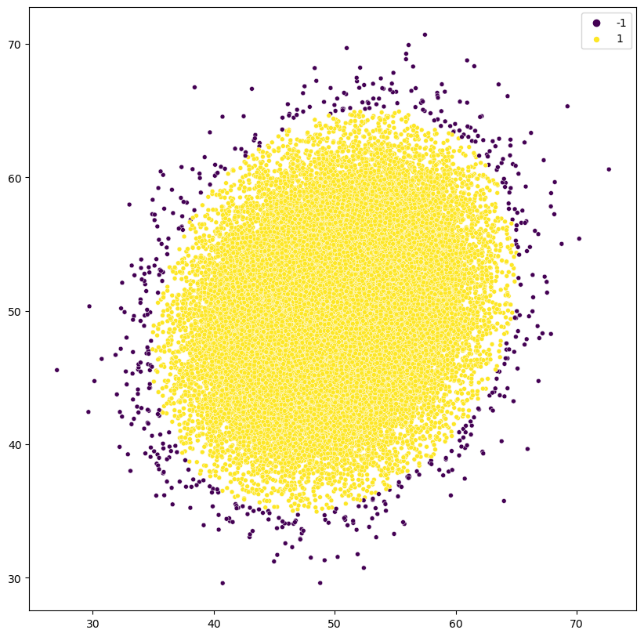
Mean :

Covariance :



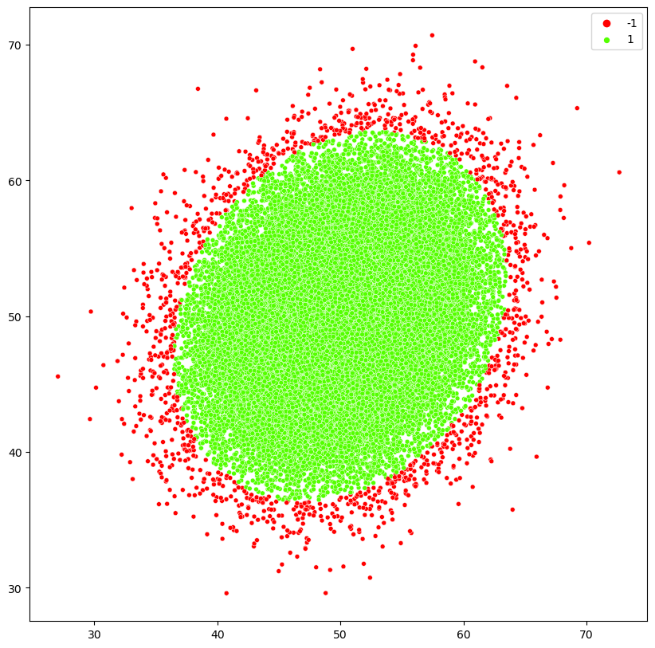
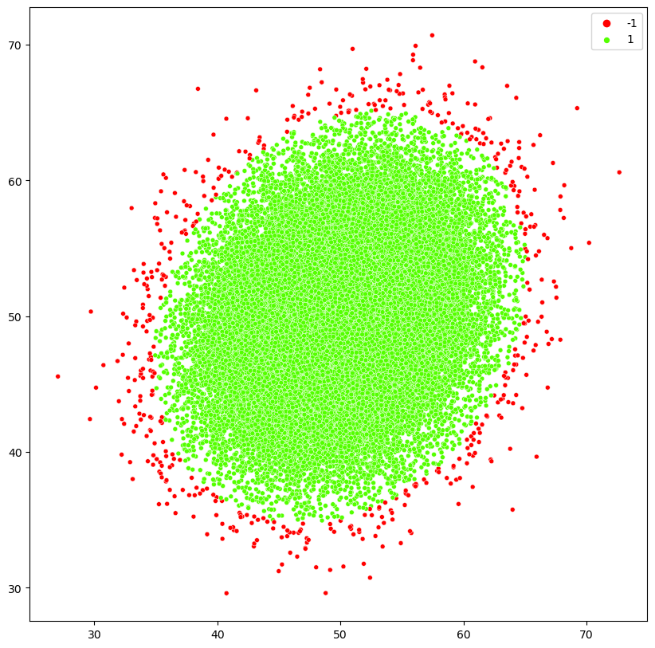
Original Data Scatter Plot

1. Mahalanobis Distance

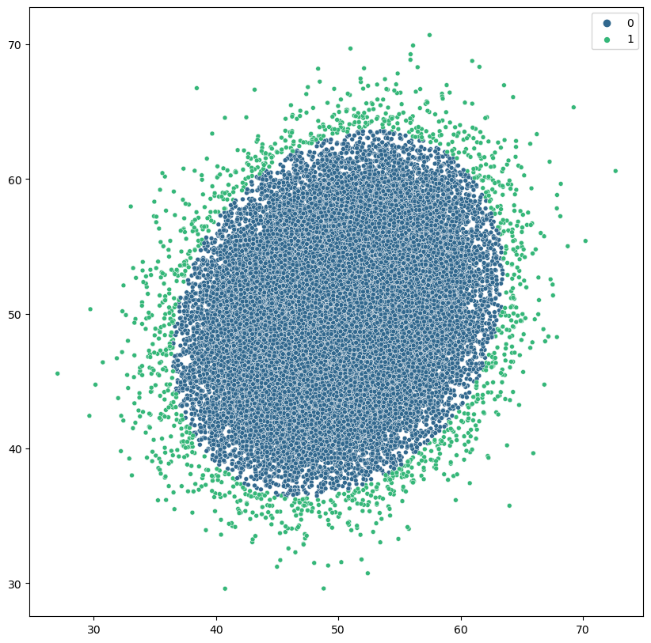
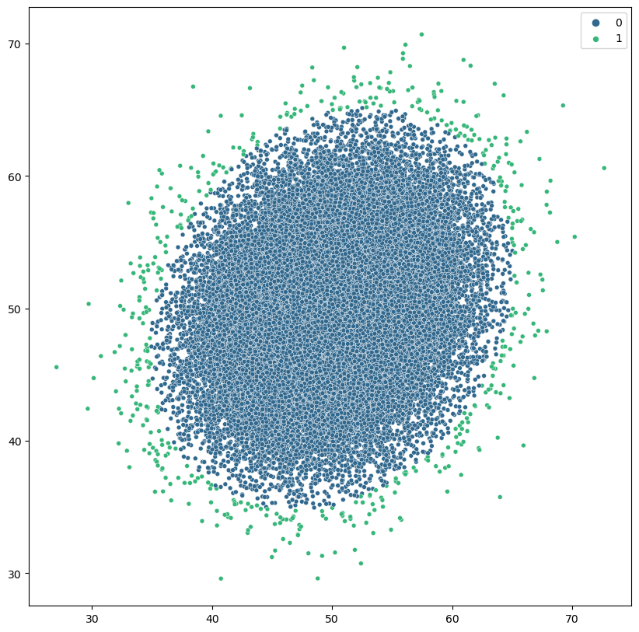
Contamination = 0.025 Contamination = 0.01

1. Elliptic Envelope

Contamination = 0.025 Contamination = 0.01

1. Minimum Covariance Determinant

Contamination = 0.025 Contamination = 0.01

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Method | Robustness to Outliers | Assumption of Data Distribution | Efficiency | Breakdown Point | Application Scenarios |
| Mahalanobis Distance | Sensitive to outliers | Assumes multivariate Gaussian distribution | Computationally expensive | Moderate | Various domains (finance, biology, engineering) |
| Elliptic Envelope | More robust than Mahalanobis | Assumes Gaussian distribution | Efficient for high-dimensional data | High | Anomaly detection, observation ranking, clustering |
| Minimum Covariance Determinant (MCD) | Highly robust against outliers | Assumes multivariate Gaussian distribution | Efficient due to fast algorithm | Very high | Robust covariance estimation in various domains |