Problem Description:
InsureNow is a leading commercial Insurance provider with PAN India presence, The company wants to build a prediction algorithm based on a subset of data from 2002 to 2009.

The company also wants to understand the data characteristics, so that it would help to negotiate better terms with the clients.

Data Description:
The dataset has 12 columns across 3 datasets:

1. Claim Number
2. Incident Date - Date of Incident
3. Date Received - Date on which claim request was received
4. City Code
5. City
6. Enterprise Type - Type of the company
7. Claim Type
8. Claim Site
9. Product Insured - Which product is being claimed for payment
10. Claim Amount - Amount in Hundreds - Amount Claimed by Client
11. Close Amount - Amount In Hundreds - Amount Approved after Inspection and Verification
12. Disposition - Target Variable - States if the Claim has been Approved, Settled or Denied

Evaluation Metric: Recall

Pre Processing Hint:
Explore and check data sanity across all columns, Eg: Is all the data between the stated time period ?

Here are a few EDA and Visualization ideas, You are required to explore more and report your findings

EDA:
1. What is the average % of claims received when the claim is 'Settled' ?
   Hint: Create a new column with % of claim received
2. Which 'Product' has the highest delay in Reporting (In days) ?
   Hint: Create a new column with Reporting delay

Visualization:

1. Do people Report Products with more value earlier compared to Products with less value?
2. Find if Delay in reporting affects the Percentage of Claim Received ?
3. How claim type and filed Claim Amount could affect the settlement of claim Amount?
4. Which City has reported the highest number of incidents ?
5. Which Type of Claims have the highest number of incidents ?

Modelling:

1. Choose a model to baseline your classification
2. Apply one or more classifiers to show an improvement on the baseline model

Model Evaluation:

For the implemented models provide the best parameters using Grid search and also present your model performance based on the given evaluation metric.