Explanation of the dataset:

No. of Columns: 15

Dependent variable: Delivery\_Status

Task: Classification

NA's: Yes

Explanation of the Columns:

Year: The Year the data was collected

Month: The Month in which the data was collected

DayofMonth: The day of the month

DayofWeek: The day of Week

Actual\_Shipment\_Time: The Actual time when the package was sent for shipment. (ex: 1955 means 19 hours and 55 minutes i.e 7:55 PM)

Planned\_Shipment\_Time: The time when the package should have been sent for shipment. (ex: 1955 means 19 hours and 55 minutes i.e 7:55 PM)

Planned\_Delivery\_Time: The time when the package should be delivered. (ex: 1955 means 19 hours and 55 minutes i.e 7:55 PM)

Carrier\_Name: The name of the Carrier which carried the package.

Carrier\_Num: The number of the Carrier which carried the package.

Planned\_TimeofTravel: The estimated time to reach from Source to Destination. ( in minutes)

Shipment\_Delay: The time by which the package was shipped late. (in minutes. Negative value indicates that the package was shipped early. Ex: 4 indicates that the package was shipped 4 minutes late, whereas, -4 indicates that the package was shipped 4 minutes early)

Source: The place from which the package was shipped.

Destination: The place at which the package was delivered.

Distance: Distance between Source and Destination in miles.

Delivery\_Status: Whether it got delivered at the right time or not. (Dependent Variable)

Expectation & Point to Remember:

CRISP-ML(Q) framework has to be implemented and a document has to be created with all sections completely filled.

Apply EDA, Feature Engineering, Data Visualization, Unsupervised Learning and finally Supervised Learning. Also take the categorical variables and perform word cloud.

Get as innovative as possible because this is the right sized dataset to get complete understanding on end-to-end pipeline building.

Accuracy of the model should be greater than 90%.

If you get 100% accuracy then something is wrong with the way you handled the dataset.

Use Google Colab to avoid running out of memory issues.