PRESENTATION

Task 5:- Modeling

Data Preprocessing (Recap)

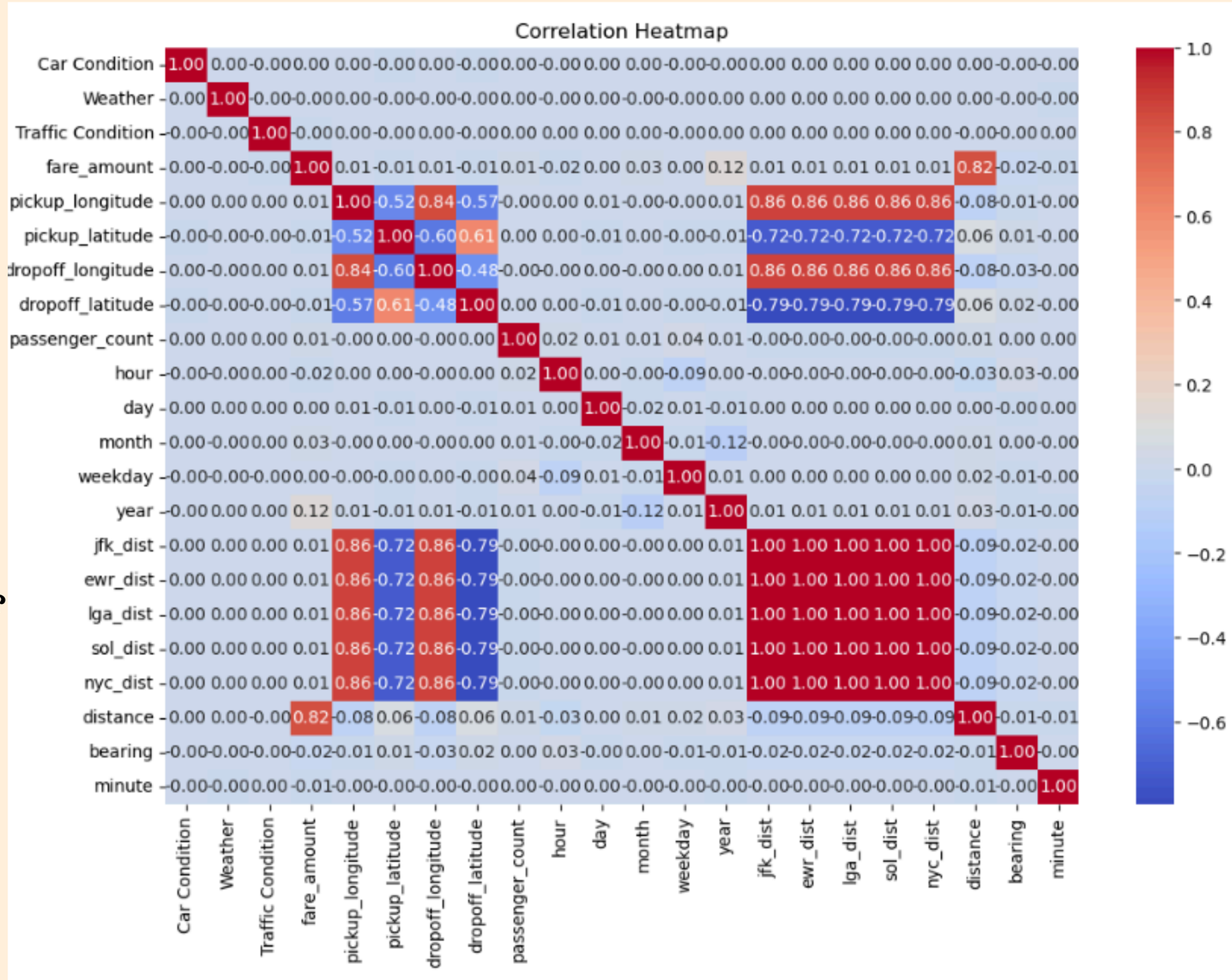
-after exploring and cleaning the dataset in part 1:-

1-check missing values.
2-check duplicated values
3-creating new features(minutes).
4-handling outliers.

Feature selection

-we analyzed the correlation between features and target (fare amount)

-we selected (distance,year,month,pickup_longitude,dropoff_longitude,dropoff_latitude)



Model 1 - linear Regrassion

-First, we trained linear regression model.

Performance:-

-mean absolute error: 2.51
-mean squared error: 28.02
-R^2 score: 0.70

Model 2 - SVC (Support Vector Classifier)

we trained SVC:-

Performance:-

-mean absolute error:2.30
-mean squared error:25.00
-R^2 score:0.73

Model 3 - Random Forest Regression

we trained random forest:-

Performance:-

-mean absolute error: 1.91
-mean squared error: 16.64
-R^2 score: 0.82

Grid Search for random forest(Hyperparameter Tuning)

-We used GridSearchCV to search for the best hyperparameters.

-Grid Search uses cross-validation to find the best combination based on lowest.

```
n_estimators: [50, 100, 200]
    max_depth: [None, 5, 10]
min_samples_split: [2, 5]
```

Best Model & Results

-Best hyperparameters.

n_estimators=200,
max_depth=10,
min_samples_split=2

-Performance after tuning.

-mean absolute error:1.80
-mean squared error: 15.20
-R^2 score: 0.85

THANK YOU