Machine Learning Exam

[Time: 2 hrs]

[Total Marks: 50]

	Perform the following tasks:	**Marks**
Q.1	Perform Feature Engineering a) Perform basic exploration like checking for top 5 records, shape, statistical info, duplicates, Null values etc. b) Extract Date, Month, Year from Date of Journey column	**[10]**
Q.2	Perform Exploratory Data Analysis (EDA) tasks a) Which airline is most preferred airline c) Find the majority of the flights take off from which source d) Find maximum flights land in which destination	**[6]**
Q.3	Compare independent features with Target feature to check the impact on price a) Which airline has the highest price b) Check if the business class flights are high price or low and find only those flights which price is higher than 50k	**[4]**
Q.4	Perform encoding for the required features according to the data	**[5]**
Q.5	Build multiple model by using different algorithm such as Linear Regression, Decision Tree, and Random Forest etc. and check the performance of your model.	**[10]**
Q.6	Compare all of the models and justify your choice about the optimum model by using different evaluation technique and tune the models as per the requirement.	**[8]**
Q.7	Write a conclusion from the business point of view. Finally perform the same preprocessing technique for test data best practice using pipeline.	**[7]**
Q.8	Calculate the a) recency (R), b) frequency (F) c) monetary value (M) for each customer based on the given dataset?	**[15]**
Q.9	a) Calculate RFM scores. Each customer will get a note between 1 and 5 for each parameter for Recency(R), Frequency(F) and Monetary value(M) Ex: Scale for Recency:	**[20]**

	1: 0-30 days 2: 31-60 days 3: 61-90 days 4: 91-180 days 5: 181-365 days b) Segment the customers based on their RFM scores using the dataset? **Segment** / **Description** Champions — Bought recently, buy often and spend the most Loyal Customers — Buy on a regular basis. Responsive to promotions. Potential Loyalist — Recent customers with average frequency. Recent Customers — Bought most recently, but not often. Promising — Recent shoppers, but haven't spent much. Customers Needing Attention — Above average recency, frequency and monetary values. May not have bought very recently though. About To Sleep — Below average recency and frequency. Will lose them if not reactivated. At Risk — Purchased often but a long time ago. Need to bring them back! Can't Lose Them — Used to purchase frequently but haven't returned for a long time. Hibernating — Last purchase was long back and low number of orders. May be lost. Segments with RFM score range: - Champions: RFM score range - R: 4-5, F: 4-5, M: 4-5 - Loyal customers: RFM score range - R: 3-5, F: 3-5, M: 3-5 - Potential loyalist: RFM score range - R: 4-5, F: 2-3, M: 2-3 - Recent customers: RFM score range - R: 4-5, F: 1-2, M: 1-2 - Promising: RFM score range - R: 4-5, F: 1-2, M: 1-2 - Needs attention: RFM score range - R: 3-5, F: 3-5, M: 3-5 - About to sleep: RFM score range - R: 1-2, F: 1-2, M: 1-2 - At risk: RFM score range - R: 2-5, F: 1-3, M: 1-3 - Can't lose them: RFM score range - R: 1-3, F: 4-5, M: 4-5 - Hibernating: RFM score range - R: 1-2, F: 1-2, M: 1-2	
Q.10	a) Visualize the RFM segments. b) Conclude your findings of RFM analysis and suggest some strategies on it.	**[15]**