1) Hasta el momento, las variables independientes de los modelos han sido valores numéricos. Investiguen el método de one hot enconding para que los modelos tomen como entrada valores categóricos. Describan este proceso y adjunten la referencia de su investigación.
2) El archivo Titanic.csv contiene información de los sobrevivientes del barco, la descripción de cada una de las variables es:

    Survived: Survived (1) or died (0); this is the target variable

    Pclass: Passenger's class (1st, 2nd or 3rd class)

    Sex: Passenger's sex

    SibSp: Number of siblings/spouses aboard

    Parch: Number of parents/children aboard

    Fare: Fare

    Embarked: Port of embarkation

En base a ese archive generen un modelo de regresión logística que determine si un paciente sobreviviría el accidente en base al resto de las variables dependientes. Hagan una predicción de si un pasajero sobreviviría en caso de ser Pclass: 1, female, Sibsp: 0, Parch: 0, Fare 7.5 y Embarked: C.

3) The dataset in q1_data.csv is comprised of 1000 instances, each with 2 features (F1 and F2) and a binary label (0 or 1). The two features are related in a particularly interesting way.

    a) Find out what this relationship looks like by plotting a 2-D graph of the two features. You can use Python's matplotlib for this task or any other plotting tool of your choice. Include the plot in your answer.

    b) Using scikit-learn, fit a logistic regression model to the dataset and evaluate its performance (accuracy) with 10-fold cross-validation. You should report performance using the accuracy measure, averaged across all cross-validation runs. Include your source code.

    c) Can you think of a way to improve the performance of the model while still employing the logistic regression algorithm? If so, describe how, include your code and present performance results. (Hint: think about creating new features based on F1 and F2).