**Dataset Description**

This dataset was retrieved from Kaggle and is used in an ongoing Kaggle competition. The competition involves predicting whether a passenger was transported to an alternate dimension during the Spaceship Titanic's collision with the spacetime anomaly. The reason why this Dataset was chosen because as an aspiring Data Scientist, I felt that I have to learn, gain the required skills and practice a lot to get more experience. And I felt that participating in data science competitions to be one of the best approaches to help beginners in data science get more experience and finally apply for job opportunities. Below mentioned are the Files retrieved and Data Field Descriptions.

**Train set**

**Filename:** train.csv

**train.csv** - A total of 8700 passengers' personal records were collected, to be used as training data.

* **PassengerId** - One unique Id for every traveler. Each Id has the following format: gggg pp, where pp is the passenger's position inside the group and gggg denotes the group they are traveling with.  People in a group are often family members, but not always.
* **HomePlanet** - The planet from whence the traveler departed, usually their home planet.
* **CryoSleep -** Whether the traveler chose to be placed in suspended animation for the duration of the trip is indicated. Cryosleeping passengers are confined to their cabins.
* **Cabin** - The passenger's assigned cabin number. , where side can be either P for Port or S for Starboard, takes the form deck/num/side.
* **Destination** - The planet from which the passenger will disembark.
* **Age -** The age of the passenger.
* **VIP -** Whether the passenger has paid for special VIP service during the voyage.
* **RoomService, FoodCourt, ShoppingMall, Spa, VRDeck** - Amount that the traveler has paid for each of the luxurious amenities on the Spaceship Titanic.
* **Name -** The first and last names of the passenger.
* **Transported -** Whether the traveller was taken to another dimension. The column we are attempting to forecast is the target, otherwise known as the target.

**Test set**

**Filename:** test.csv

**test.csv -** The personal information of the remaining 4300 passengers will be used as test data.

**Methodology**

In this project report, we will cover the following:

* Exploratory Data Analysis
* Feature Engineering
* Data Cleaning
* Encoding, Scaling and Pre-processing
* Training Machine Learning Models
* Cross Validation and Ensemble Predictions