**Data Description**

In this competition your task is to predict whether a passenger was transported to an alternate dimension during the *Spaceship Titanic*'s collision with the spacetime anomaly. To help you make these predictions, you're given a set of personal records recovered from the ship's damaged computer system.

File and Data Field Descriptions

* train.csv - Personal records for about two-thirds (~8700) of the passengers, to be used as training data.
  + PassengerId - A unique Id for each passenger. Each Id takes the form gggg\_pp where gggg indicates a group the passenger is travelling with and pp is their number within the group. People in a group are often family members, but not always.
  + HomePlanet - The planet the passenger departed from, typically their planet of permanent residence.
  + CryoSleep - Indicates whether the passenger elected to be put into suspended animation for the duration of the voyage. Passengers in cryosleep are confined to their cabins.
  + Cabin - The cabin number where the passenger is staying. Takes the form deck/num/side, where side can be either P for *Port* or S for *Starboard*.
  + Destination - The planet the passenger will be debarking to.
  + 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 the passenger has billed at each of the *Spaceship Titanic*'s many luxury amenities.
  + Name - The first and last names of the passenger.
  + Transported - Whether the passenger was transported to another dimension. This is the target, the column you are trying to predict.
* test.csv - Personal records for the remaining one-third (~4300) of the passengers, to be used as test data. Your task is to predict the value of Transported for the passengers in this set.
* sample\_submission.csv - A submission file in the correct format.
  + PassengerId - Id for each passenger in the test set.
  + Transported - The target. For each passenger, predict either True or False.

### Dataset Link:- https://www.kaggle.com/competitions/spaceship-titanic/data

### Objective

\* In this notebook we will focus towards conducting a thorough exploratory data analysis to understand the data and the relationship between the variables.

\* We will also perform feature engineering to create new features that are informative.

\* After that we will fit a Random Forest model to create predictions and submit.

### Problem Summary

\* We are in the year 2912 and a spaceship called Titanic started off from the Earth to emigrate about 13,000 passengers to three newly habitable exoplanets.

\* Before reaching it's first destinantion it collided with a spacetime anomaly and almost half of the passengers were transported to an alternate dimension.

\* Using the records recovered from the spaceship, we need to predict which passengers were transported by the anomaly.

\* Link to the competition.

\* Table of Contents

1. Import packages & Read data

2. Utility Code

3. Inspect the data

4. Exploratory Data Analysis

4.1 Transported

5. Feature Engineering

6. Random Forest Model

6.1 Data Preprocessing

6.2 Fitting a Random Forest Model

7. Submission

8. References