**Market Basket Analysis on Survival of Titanic Passenger**

Data Description:

* Survival: Survive or not (0: No, 1: Yes)
* Pclass: Ticket Class (**1**: 1st, **2:** 2nd, **3** :3rd class)
* Sex: Male or Female
* Age: Age in years
* Sibsp: No. of siblings(**brother, sister, stepbrother, stepsister** ) /spouses(**husband, wife**) around the titanic
* Parch: family relations(**Parent** = mother, father/ **Child** = daughter, son, stepdaughter, stepson  
   / Some children travelled only with a **nanny**, therefore parch=0 for them
* Ticket: Ticket number
* Fare: Passenger Fare
* Cabin: Cabin Number
* Embarked: Port of Embarkation.( C:Cherbourg, Q: Queenstown, S: Southampton)

Our Dataset is divided into two groups training and testing dataset. Training set has 891 rows and 12 columns and test set has 418 rows and 11 columns. Also, test dataset has no “Survival” Column. We’ll strip out the “Survival” Column from train dataset and combine train and test data.

Since Market Basket Analysis is used with only logical (i.e., binary) variables.

So, I have added some more binary features: “**Firstclass**” , “**SecondClass**”, “**ThirdClass**” having Pclass value 1, 2 and 3 respectively . “**Female**” features : whether female or not, “**AgeMissing**”: as missing value of age or not, “**Child**”: having age 0 to 17 or not, “**Adult**” : having age 17 to 55 or not , “**Elderly**”: having age greater than 55 or not .

Now storing these 19 columns with rows in a new data frame “**titanic.features**”.

Dataset has a feature “**Ticket**” represents the number of passenger allowed for that ticket. I use this information and create some more binary features : “**IsSolo**”, “**IsCouple**”, “**IsTriplet**”, “**IsGroup**” for having ticket only for 1person,2 persons , 3 persons and more than 3 persons respectively.

We also create “**HasChild**”, “**HasElderly**” for those whose ticket contain information of age 0 to 18 and of age greater than 55 respectively and “**Noages**” for those tickets having age feature null .

I now store these 8 columns with rows in a new Dataframe “**titanic.groups**”.

I now combine all the logical features in a new dataset “**titanic.binary**” which contains 15 columns .(‘ticket’

column which is not binary is removed )

Using Aprori function

1. Join “titanic.features” and “titanic.groups” using inner join function.
2. Select only binary features.

Now Separating the training set and adding back the “Survived” Column to a new Dataframe **“ train.binary**”.

Assigning 0,1 in Survived column where 1: true, 0: False

Market Basket Analysis:

It consists of mining frequent associations from the data which are referred as Association Rule or “Association Rule Mining”.

Transforming the “train.binary “ dataframe to transaction objects dataframe named “**titanic.trans**”.

Rules of Aprori:

For single feature:

1. Specify minimum support of 0.05 and confidence of 0.1.
2. Specify a maximum of two “items”.(eg Female 🡪 Survived)
3. Get those “items” who are highly associated with “Survival” features.
4. Get top 20 mined rules in term of “Lift Metric”.
5. Pinned “Survived” variable on right hand side and “Association Rules” in left hand side.

For Multi features:

1. Specify minimum support of 0.05 and confidence of 0.1.
2. Specify a maximum of four “items”.
3. Get those “items” who are highly associated with “Survival” features.
4. Get top 20 mined rules in term of “Lift Metric”.
5. Pinned “Survived” variable on right hand side and “Association Rules” in left hand side.

Observations:

For first Association Rule {**FirstClass, Female, Adult**} => {**Survived**}

1. The “support” metric tells what percentage of overall “transactions (i.e. passengers)” exhibit the rule. In this case, 7.63% of all passengers.
2. The “Confidence” metric tells what percentage of passengers that fit the rule on the left-hand side also had the right-hand side (i.e., Survived). In this case, 97.14%.
3. The lift metric tells how many more times likely is the right-hand side to happen, assuming the left-hand side, compare to all passengers. In this case, Survived is 2.53 times more likely.

Conclusion:

“**Adult females** travelling in **first class** account for 7.63% of titanic passengers. These passengers survived at a rate of 97.14%. Also, a first class adult female passenger were 2.53 more likely to survive than a random passenger on the Titanic.