# Spaceship Titanic Project

On my .Rmd file with all my queries I wrote a lot of analysis on what I was thinking at the time, so on the following pages I'm just going to summarize some of my findings and highlight what I found most interesting as I was analyzing the Spaceship Titanic dataset. First off, I think it's safe to say we have somebody out in the future with a nefarious plan to kidnap Martian children all under the guise of a "space-time anomaly". Too bold of an accusation? Probably so; while fun to think about an intergalactic conspiracy there was nothing conclusive to back up that claim. Which brings me to my point, while finding some interesting characteristics in the data, without more information I'm not able to definitively say what the reasoning is behind a passenger being transported or not. To be more specific, I need a precise layout of the ship to take analysis any further or at least without the help of machine learning tools. I could be overlooking something but I was having trouble figuring out floorplan of decks (A,B,C, etc.) and levels(0-1500+) in relation to each other and without knowing for certain their positioning on ship I would just be making guesses in analysis and would have high probability of not being accurate.

There were some interesting trends in the data though and I'll touch on them now. Also in the following pages anytime you see TPR, it stands for Transportation Rate, a stat that was very important to my analysis. TPR represents the probability of transportation among certain subgroups.

Top Findings:

- **Cabin (Deck)** – The Cabin field was split into three characteristics, Deck/Level/Side (i.e. A/145/P), and was a key area of my analysis. Whether a passenger was on the Port or Starboard side didn't seem to have much impact on whether a passenger was transported. In contrast
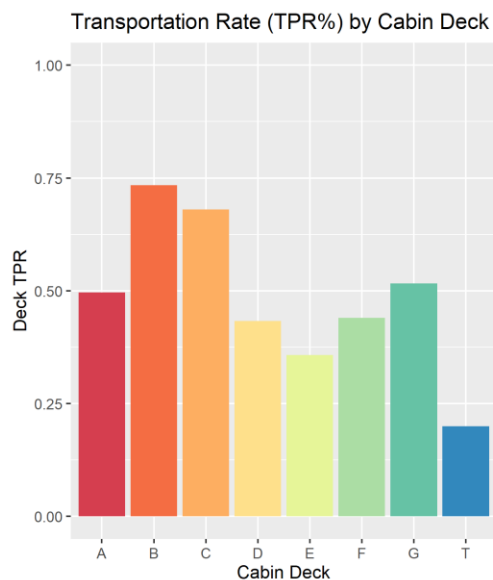
examining the passenger's deck did yield some interesting points. Shown in the graph to the left, you will notice the high TPR's in Decks B and C in comparison to most others that hover around 50%. You can disregard Deck T, the amount of passengers on that deck is negligible compared to the others so it shouldn't of been included in graph. A few more numbers to give you better grasp of graph, the sample size of Decks B and C was roughly 1,500 passengers, they were the 3rd and 4th most populated decks. Decks F and G was where the bulk of passengers were located though, with over 5,000 passengers staying on their decks (over 50% of entire train.csv file).

- **Cabin (Level)** – I thought Cabin Level was most interesting out of the three characteristics. While analyzing passengers who went missing on each level I noticed a trend shown in graph below. Starting at level 4 you begin to see a slight increase in TPR on each level reaching a peak at level 9 and then it begins decreasing again until you reach level 14 (note: levels 15+ were disregarded due to number of passengers staying on that level being too small a sample size).

Level TPR

TPR values by Cabin Level: 53%, 54%, 54%, 43%, 36%, 43%, 50%, 54%, 58%, 67%, 63%, 51%, 44%, 42%
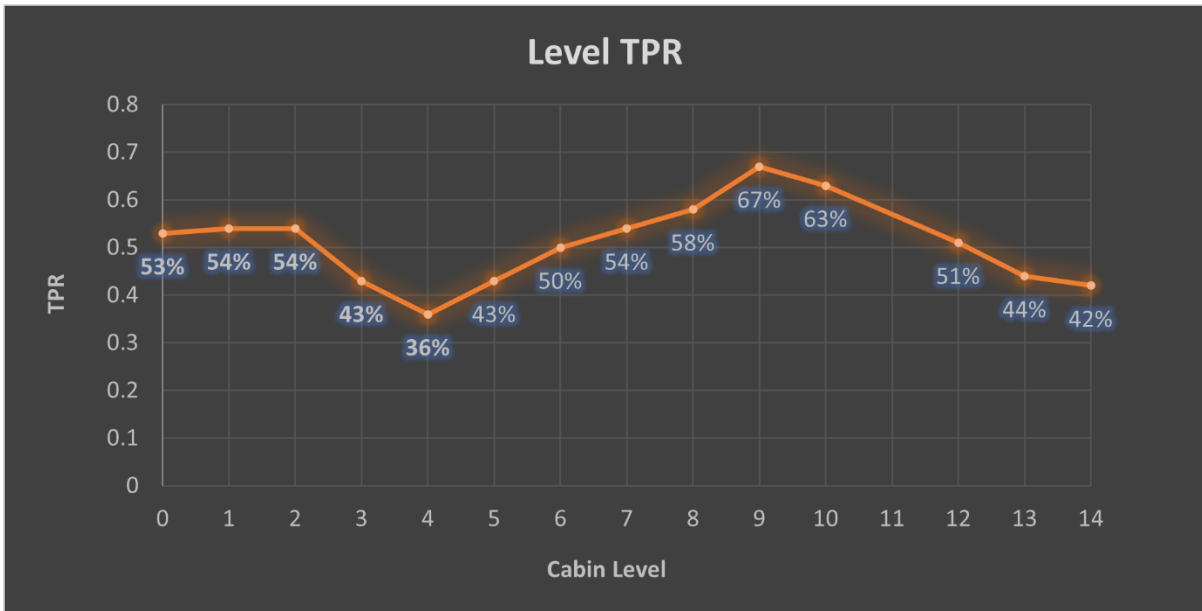
When this is mapped out in a graph you see a bell curve distribution among levels 4-14, which I found very interesting. The outliers being the first few levels (0-3), that show relatively high TPR's compared to the rest of the levels and taken into account that most of the passengers were staying on the first few levels it was too large a sample size to be ignored. Still very interesting data, why did levels 4-14 (which had similar passenger counts on their levels) exhibit the TPR pattern that they did, there was 30% difference between levels 4 and 9 and then another 25% drop down to level 14. This is where I would have loved a map of the floorplan so I could get a better understanding of layout of the cabins.

- **Children Taken at High Rates** - Another area that was very interesting was looking into the TPR among different ages. I touched on in .Rmd file but newborns (passengers under 1 yr old) were taken at 81% rate with children under the age of 12 having an average TPR hovering around 70%. While I found it interesting, I didn't think it had to do with their age but maybe with their location on the ship. If children were all grouped together in say a daycare facility or families were placed on similar decks or levels and the space-time anomaly affected that part of the ship then that could explain why the TPR was so high for children. While looking into how many children were located one each deck I noticed a scary trend for kids that were on Decks A-F, there was a 95% average TPR, which translates to almost all the kids on those decks were transported to an alternate dimension, you can see this below in the graph on the left side. I found it really interesting but the sample size is also fairly small so I didn't think it was smart to use it as a basis for anything at least until I had more information.

## Cabin Deck TPR - Children 12 & Under
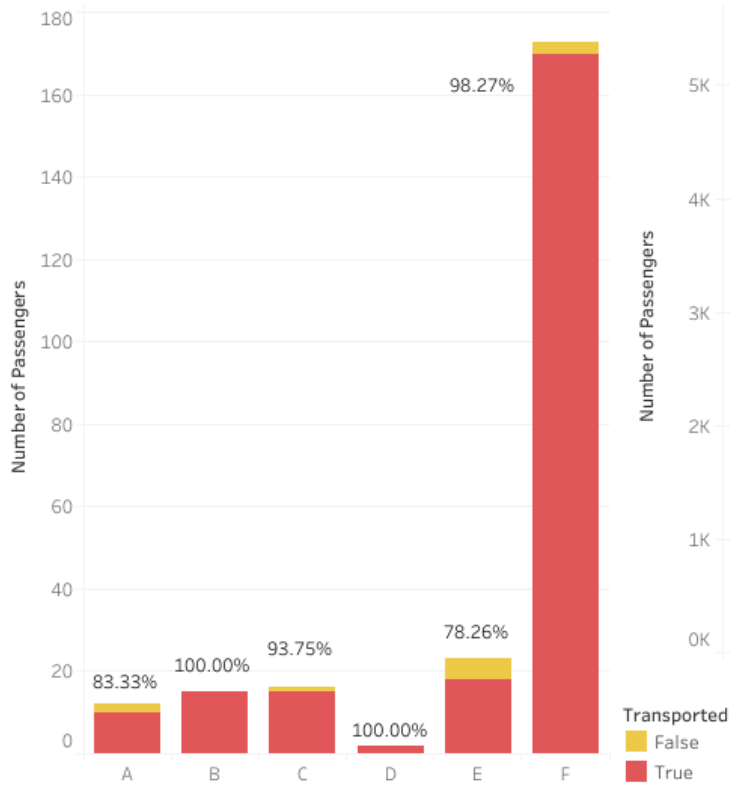
CryoSleep TPR
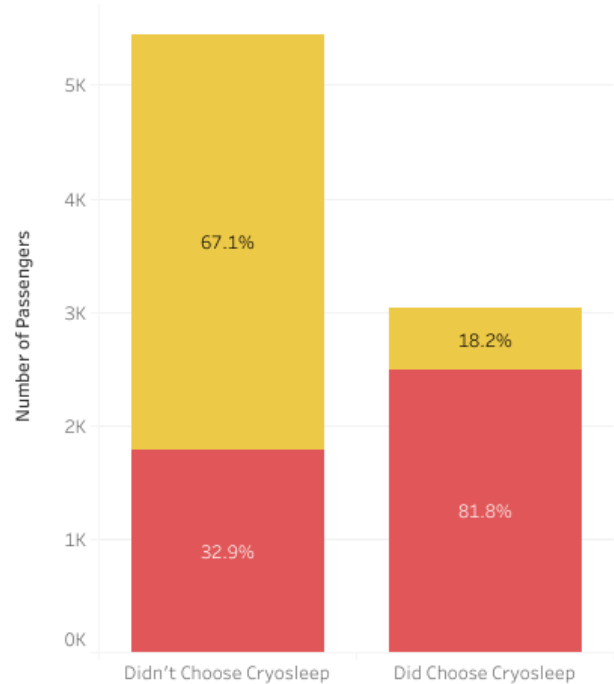
- **Cryosleep Passengers** – So out of all my analysis I think this is the most important finding, out of the roughly 3,000 passengers that choose the cryosleep option, 82% of them were transported. This is where knowing layout of ship would have been handy, since cryosleep passengers were contained to their cabins for the journey I could've been able to generate a map of affected areas if I knew where the cabins were located. You can see the relationship of cryosleepers and non cryosleepers TPR above in graph on the right side.

Summing up, I think I found some interesting trends but I think it would be irresponsible of me to make any grand claims on why passengers were transported and others not with the evidence I have. I'm curious what the solution is!
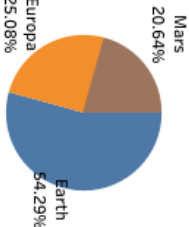
*Also included below is a graph representing age breakdown of passengers and their respective TPR's; as well as the bit of information that made me joke about the Martian children kidnapping conspiracy!

Starship Titanic

Transportation Rate (TPR) by Age
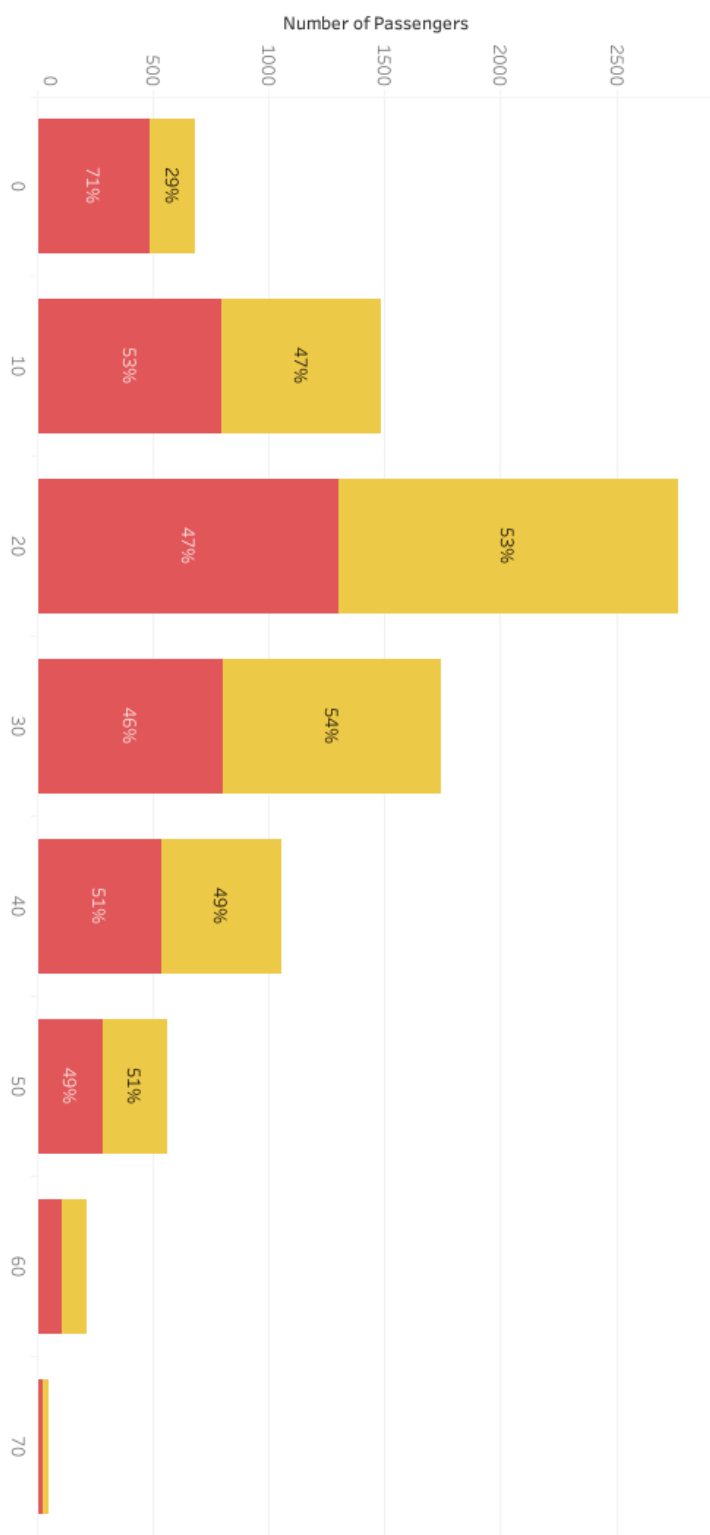
Home Planet Breakdown - Entire Ship

Mars 20.64%
Europa 25.08%
Earth 54.29%

Home Planet Breakdown - Missing Children 12 & Under

Mars 74%
Earth 9%
Europa 17%