*“We had no time to think now, only to act. We shook hands, wished each other luck. I said, ‘Go ahead, I’ll be right with you.’ I threw my overcoat off as he climbed over the rail, sliding down facing the ship. Ten seconds later I sat on the rail. I faced out, and with a push of my arms and hands, jumped into the water as far out from the ship as I could. When we jumped we were only 12 or 15 feet above the water.*

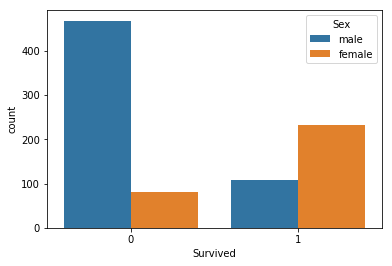
*I never saw Long again. His body was later recovered. I am afraid that the few seconds elapsing between our going, meant the difference between being sucked into the deck below, as I believe he was, or pushed out by the backwash. I was pushed out and then sucked down.”* – Jack Thayer, Titanic Survivor.

A little before midnight, the RMS Titanic crashed into an iceberg, “creating a series of holes below the waterline.” Despite being billed as the world’s largest and most luxurious ocean liner, the water started filling each compartment of the Titanic quickly. Passengers and crew soon realized how woefully unprepared the Titanic was for such a catastrophic event. The ship carried 2,208 passengers on her maiden voyage but only brought enough lifeboat seats for 1,178 passengers due to the belief that ships as large as the Titanic were “unsinkable.”

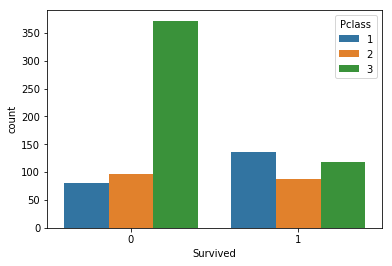
Jake Thayer, a male first-class passenger, survived the shipwreck - so did Dr. Washington Dodge, another male first-class passenger. Dr. Dodge later wrote in a letter, “One lad of 10 Master Carter told me that after his mother & sister were in a boat he was refused permission, [inserted: to enter it] and Col Astor who was standing by & who knew the lad a moment later picked up a girls hat & placing it on the lads head lifted him up & said to the officer as he was lowering the boat let this little girl go with her mother – As the lad expressed it ‘they did not stop to examine me so I got with mother’”

The three of them survived but 1,503 weren’t as lucky. Jake Thayer and Dr. Washington Dodge were both first class passengers whereas the third male pretended to be a girl in order to survive. From these anecdotes, we can infer that passengers in first class by in large survived at a higher percentage than other classes presumably due to their wealth, stature, or resources. Similarly, we can further infer that women survived at a higher percentage due to the preference that women and children were to be evacuated first.

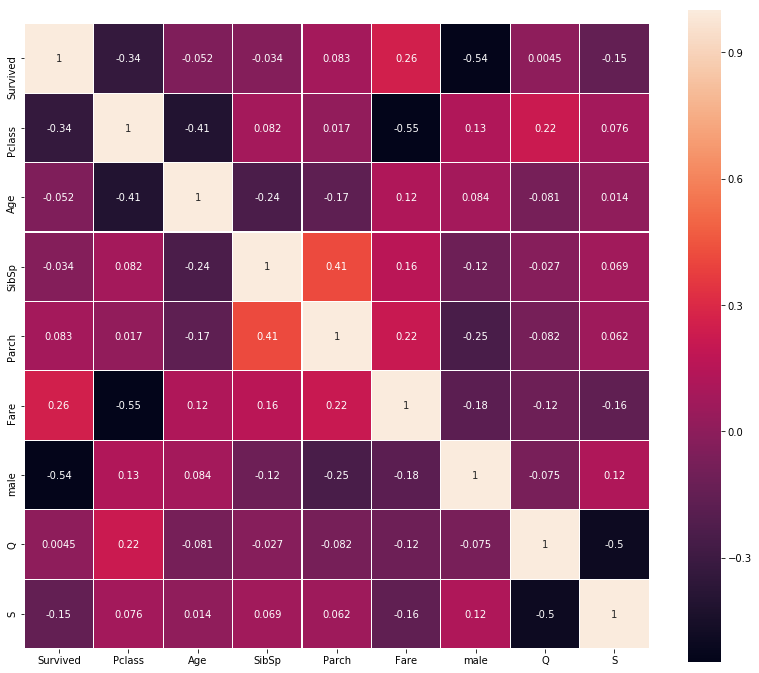
Does the data back this up?



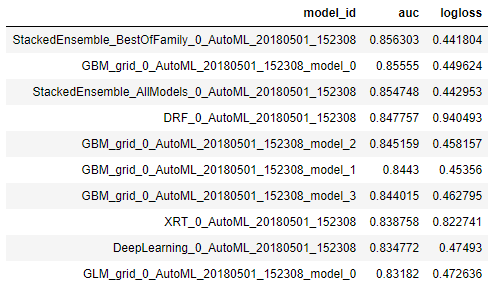
Here, with 0 being “Did Not Survive” and 1 being “Survived,” we can see that an overwhelming number of passengers who perished were men. On the flip side, the number of women who survived was more than doubled that of men.



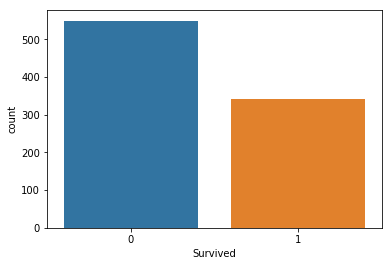
When looking at passenger classes, the highest population of people who perished came from passenger class three. However, when looking at the survivors, while first class had the most survivors it’s not a sizeable difference from second or third class.

When we look at the entire dataset we see that gender was the largest correlating factor with survivorship with fare and passenger class following that.

This begs the question – Can we predict whether or not passengers like Jake and Dr. Dodge survive? To help with this, I took a look at H2O, an AutoML product by H2O.ai. You can see all the nitty gritty details in the linked Github page but here is the finished leaderboard:



This leaderboard is evaluated by the highest AUC (Area Under Curve). AUC allows us to understand how well a classifier predicts in the test set with 0.5 being a random guess and 1 being a perfect classifier. The reason we are using ROC/AUC scores instead of simple accuracy is that there is a class imbalance in this dataset with more people having perished than having survived.



For example – let’s say you have 9,990 pairs of non-Nike shoes and 10 pairs of Nike shoes. You want to build a classifier that will predict if a shoe is a Nike shoe or not. But instead of building a real classifier, you build one that will predict “Not Nike” every single shoe. By looking at simple accuracy, you will have 99.9% model accuracy despite it getting ever single actual Nike shoe incorrect. Because there’s a class imbalance here (9,990 vs 10 shoes), it is best to use other metrics such as ROC/AUC.

In this case, the highest AUC derived from H2O was 85.63%. All in all, we have a pretty good model.

Sources:

1. <https://www.nytimes.com/1997/04/08/science/toppling-theories-scientists-find-6-slits-not-big-gash-sank-titanic.html?pagewanted=all>
2. <https://www.kaggle.com/c/titanic>
3. <https://abisaab.wordpress.com/2012/04/30/17-year-old-titanic-survivor-chilling-account/>
4. <https://www.gilderlehrman.org/sites/default/files/inline-pdfs/t-07640_0.pdf>
5. <https://www.youtube.com/watch?v=21Igj5Pr6u4&t=155s>