

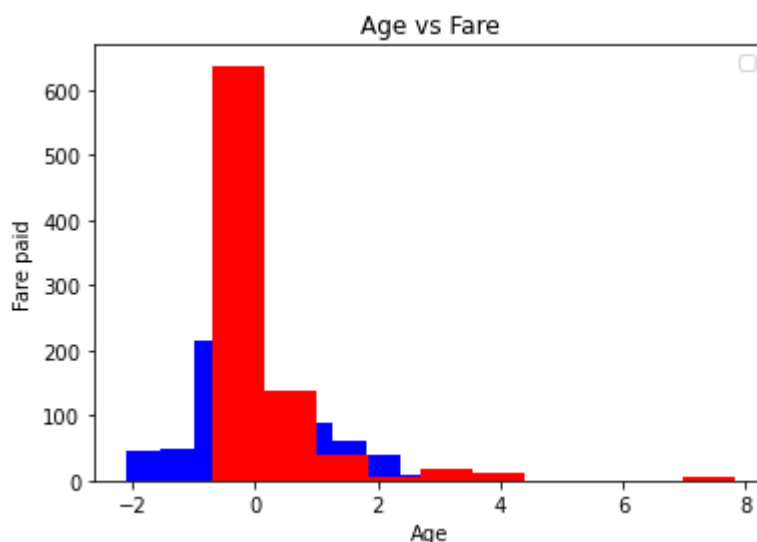
TITANIC PROJECT

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29.00	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0.92	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2.00	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
3	1	0	Allison, Mr. Hudson Joshua Creighton	male	30.00	1	2	113781	151.5500	C22 C26	S	NaN	135.0	Montreal, PQ / Chesterville, ON
4	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.00	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
5	1	1	Anderson, Mr. Harry	male	48.00	0	0	19952	26.5500	E12	S	3	NaN	New York, NY
6	1	1	Andrews, Miss. Kornelia Theodosia	female	63.00	1	0	13502	77.9583	D7	S	10	NaN	Hudson, NY
7	1	0	Andrews, Mr. Thomas Jr	male	39.00	0	0	112050	0.0000	A36	S	NaN	NaN	Belfast, NI

The dataset

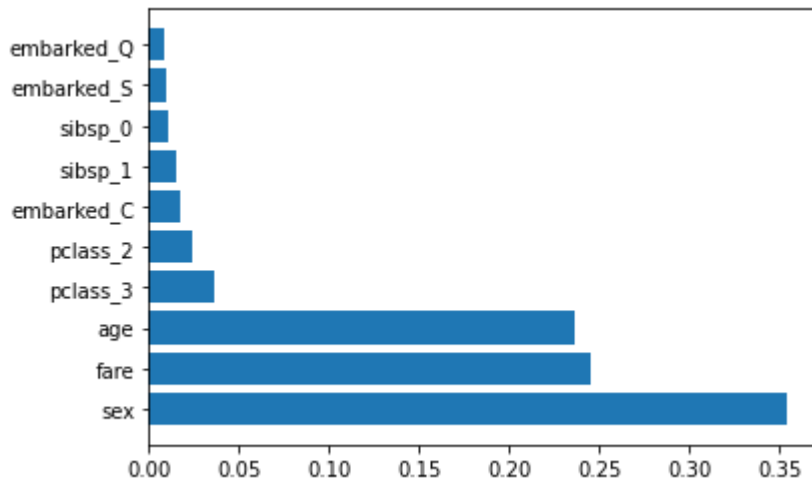
The Titanic Dataset contains data on 1300+ people who were aboard the titanic. For each of these people, various attributes such as their name, sex, age, and cabin number were recorded. The goal of this project was to explore and identify the factors correlated with surviving the crash.

I first dropped columns unrelated to survival. I then scaled the data using StandardScaler before analyzing it. StandardScaler removes the mean and scales each variable to unit variance (variance of 1).



People who paid the most amount of money were of average age - in other words not too old nor too young. People who were the youngest and the oldest paid the least. The farther you are from the average age, the less you paid to get on the ship. This could indicate that there was a discount for younger people and senior citizens. It could also indicate that the youngest and the oldest people on board did not have much money to spend, possibly because they do not have a job. It could also indicate that the youngest and the oldest

people are unwilling to spend a lot of money aboard the ship. This could be for a variety of reasons, such as they may be saving up, but falls under the category of personal preferences.



The three most important factors in determining whether a person survived the Titanic crash or not were gender, amount paid and their age respectively.

```
In [47]: #Random forest score
rf.score(test_x, test_y)
```

```
Out[47]: 0.7558139534883721
```

```
In [48]: #Decision tree score
dtc.score(test_x, test_y)
```

```
Out[48]: 0.7325581395348837
```

```
In [49]: #Logistic regression score
r.score(test_x, test_y)
```

```
Out[49]: 0.3611857325400025
```

The random forest model was the most accurate model in predicting whether a person would survive the titanic crash or not at 75.58%. The decision tree model was the second most accurate model in predicting whether a person would survive the titanic crash or not at 73.26%. The logistic regression model was the least accurate model in predicting whether a person would survive the titanic crash or not at 36.11%.