

My Sinking Capstone

DSI Capstone
By Nora Cook
April 9, 2018

Black Friday Sales

- Originally I was going to do Black Friday a Sales Competition
- Interesting dataset
- Super clean with no null values
- Turns out it was rigged
- All age ranges of buyers had almost equal mean spending.
- 0-17 year olds don't have as much money as older age ranges

No Capstone Project

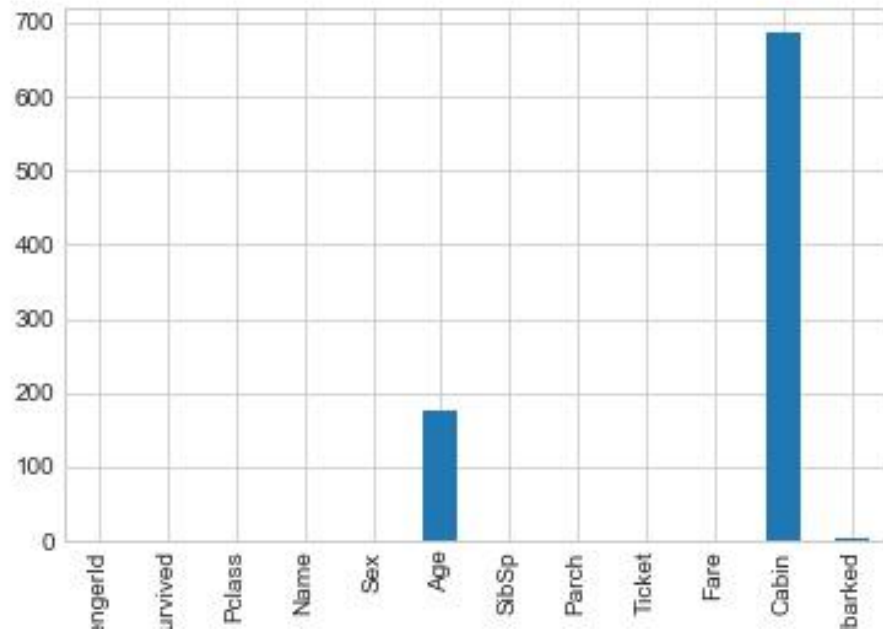
What to do?

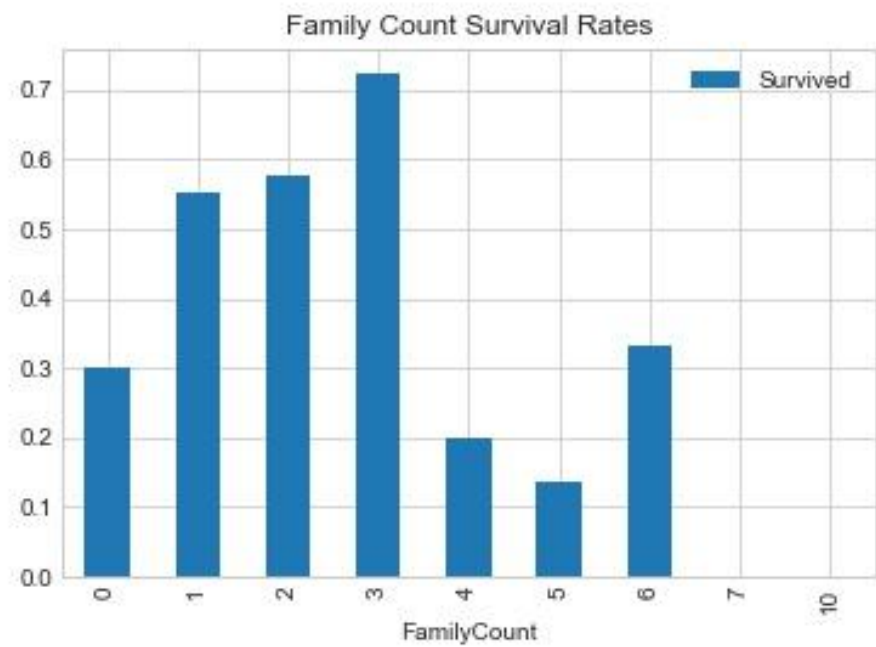
The Titanic Revisited

Problem Statement

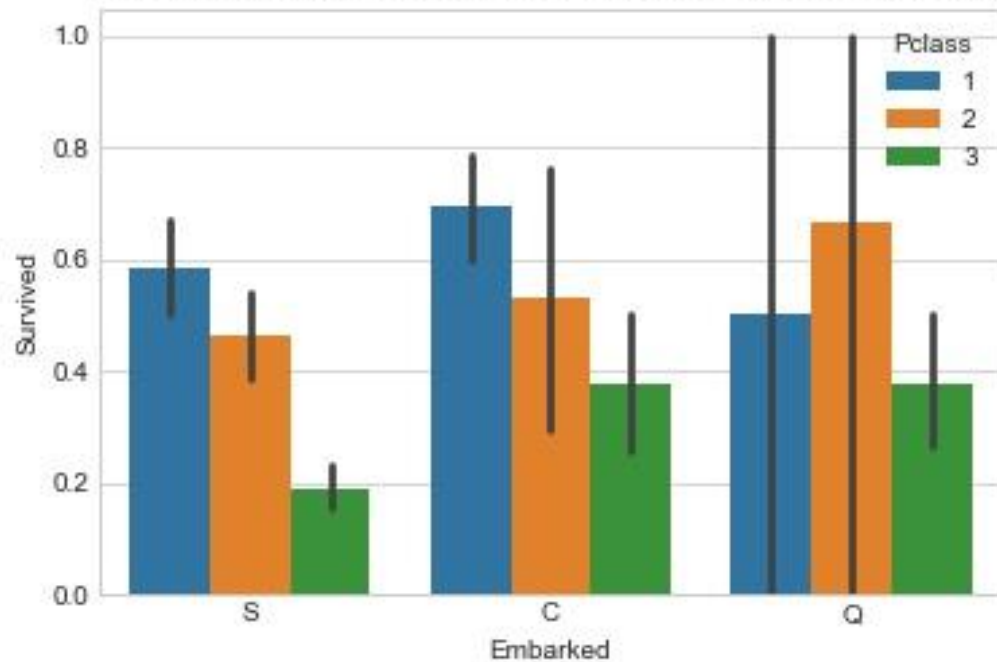
Like the state of my project after having to change data sets at the last minute the titanic was a disaster. For those that have forgotten the tale and the movie on April 15, 1912, the Titanic collided with an iceberg killing 1502 out of 2224 passengers. Recently, the Titanic tale has become a thing of beauty for Data Scientist trying to hone their machine learning skills.

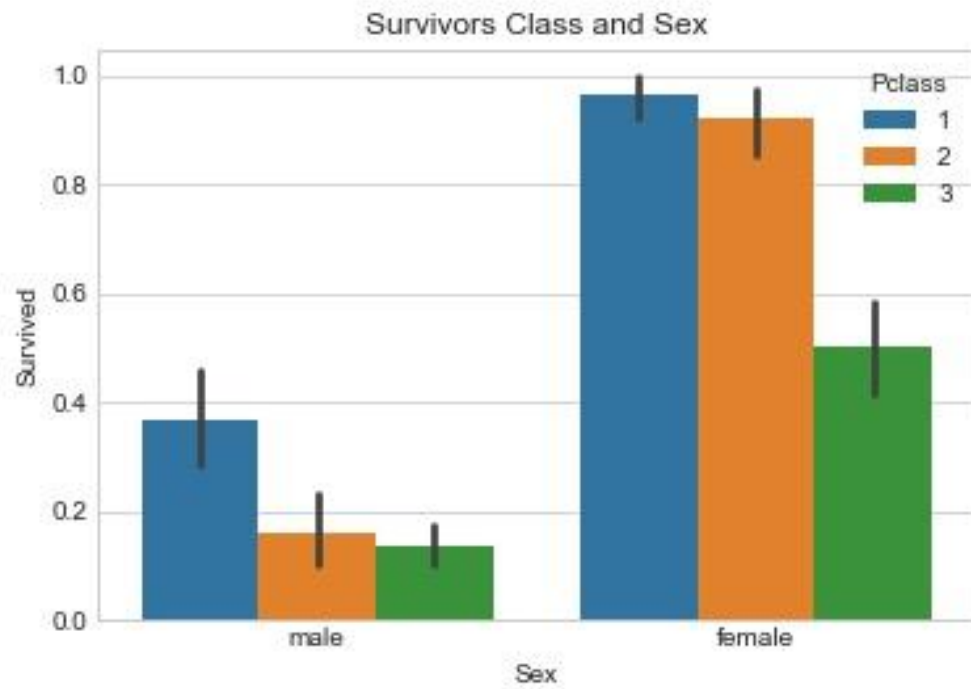
Null Values

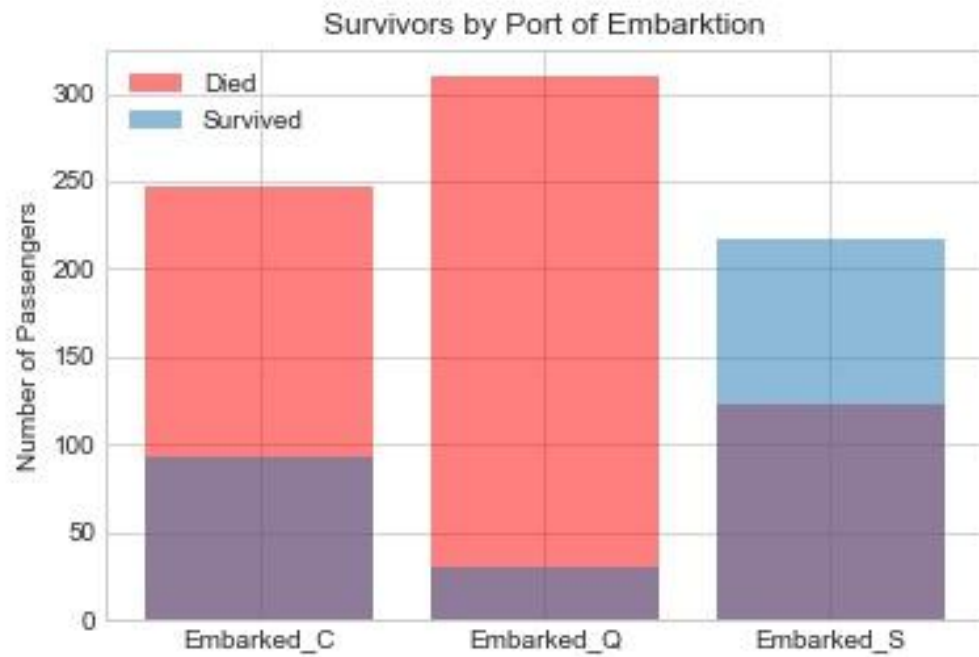




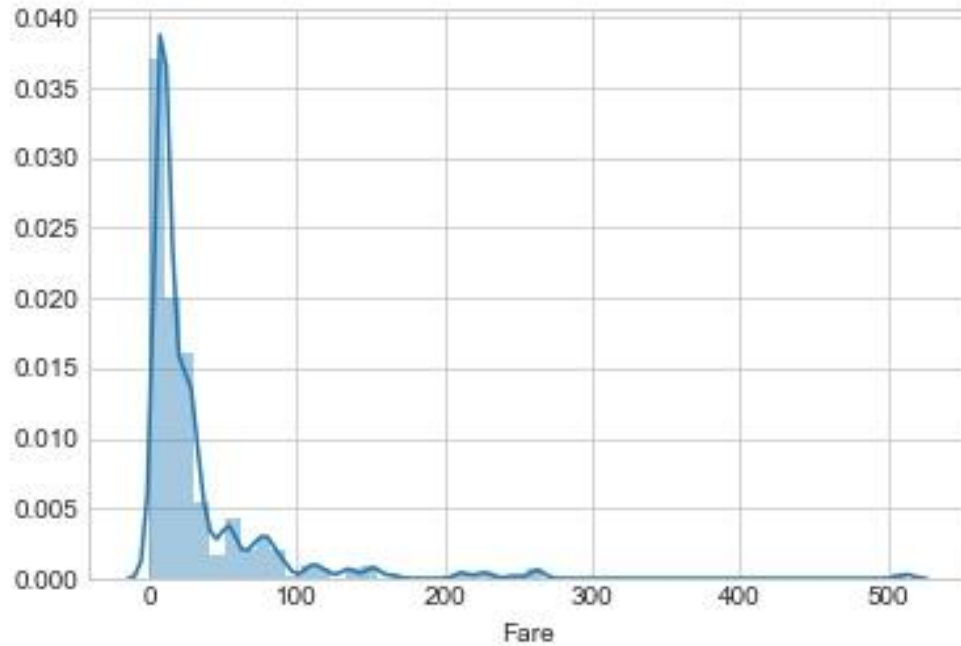
Survivors Depending on Where They Embarked and Their Class



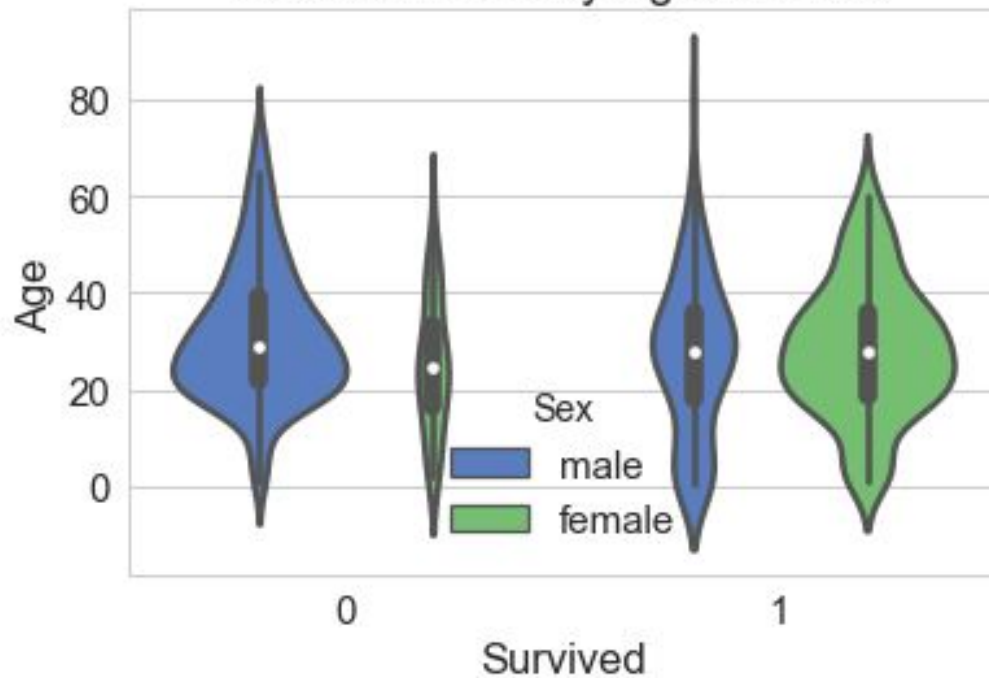




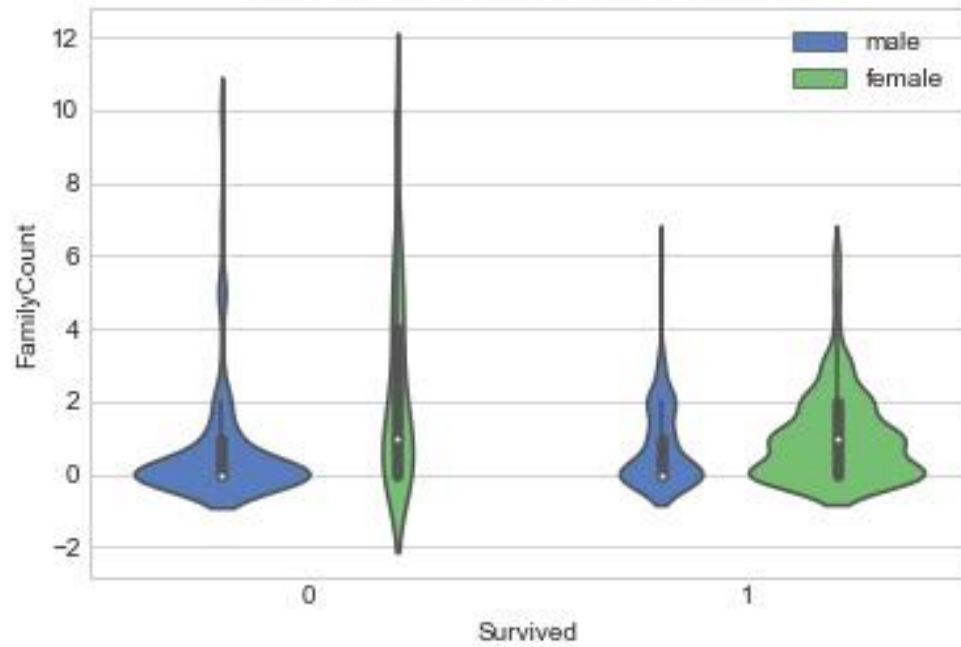
Fare Distribution



Survival Rates by Age and Sex

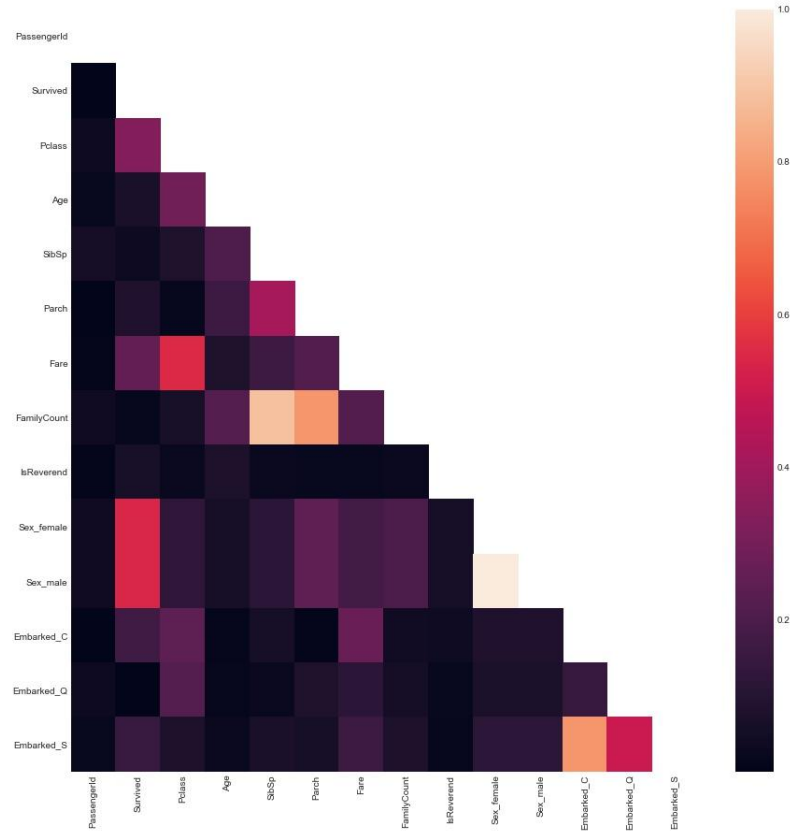


Survival Rates by Age and Family Count



What Columns did I keep

- PassengerId
- Pclass
- Age
- Sibling Spouse
- Parent Child
- Fare
- Family Count
- Sex
- Port of embarkation



Titanic Passenger Data Analysis - Feature Importance														
	Passenger Id	Survive d	Pclass	Age	SibSp	Parch	Fare	FamilyCo unt	IsRevere nd	Sex_fema le	Sex_ma le	Embarked _C	Embarke d_Q	Emba rked_ S
Passenger Id	1.000000	-0.005028	-0.035330	0.024020	-0.057686	-0.001657	0.012703	-0.040258	0.012730	-0.043136	0.043136	-0.001208	-0.033694	0.022269
Survived	-0.005028	1.000000	-0.335549	-0.069907	-0.034040	0.083151	0.255290	0.018277	-0.064871	0.541585	-0.541585	0.169966	0.004536	-0.151777
Pclass	-0.035330	-0.335549	1.000000	-0.287864	0.081656	0.016824	-0.548193	0.064221	-0.030788	-0.127741	0.127741	-0.245733	0.220558	0.076466
Age	0.024020	-0.069907	-0.287864	1.000000	-0.205270	-0.164226	0.080703	-0.222362	0.077389	-0.063014	0.063014	0.014780	0.019489	-0.025218
SibSp	-0.057686	-0.034040	0.081656	-0.205270	1.000000	0.414542	0.160887	0.890654	-0.026717	0.116348	-0.116348	-0.060074	-0.026692	0.069438
Parch	-0.001657	0.083151	0.016824	-0.164226	0.414542	1.000000	0.217532	0.782988	-0.022061	0.247508	-0.247508	-0.011588	-0.081585	0.061512
Fare	0.012703	0.255290	-0.548193	0.080703	0.160887	0.217532	1.000000	0.218658	-0.022876	0.179958	-0.179958	0.270731	-0.116684	-0.163758
FamilyCou nt	-0.040258	0.018277	0.064221	-0.222362	0.890654	0.782988	0.218658	1.000000	-0.029284	0.203191	-0.203191	-0.046852	-0.059007	0.078197

K Neighbors Classifier

Cross Validation Score = 0.792

Grid Search K Neighbors Classifier

`N_neighbors = 4`





`'P' = 1`

`weight=uniform`

Best score on model =.802

Gs.score on test=.825

K Neighbors Classifier Kaggle Submission

7925	▲ 1190	Cedric Zhao		0.76555	10	31m
7926	new	Nora Cook		0.76555	1	~10s
Your Best Entry ↑ Your submission scored 0.76555, which is not an improvement of your best score. Keep trying!						
7927	▼ 843	ghostband		0.76076	1	2mo
7928	▼ 843	XinyuChang		0.76076	3	2mo

0.76555

Random Forest

MaxDepth =None

Max Features = log2

N Estimators = 20

Gs_rf.score model= 0.817

Gs.score on test= 0.825

Random Forest Kaggle Submission

Name	Submitted	Wait time	Execution time	Score
submission_rf4.csv	12 hours ago	0 seconds	0 seconds	0.72248

Complete

[Jump to your position on the leaderboard ▼](#)

0.72248

Gradient Boosting Classifier

Maxdepth = 2

N Estimators = 50

Best Score 0.827

Score on Holdout 0.839

Gradient Boosting Classifier Kaggle Submission

5218	new	Nora Cook		0.77990	10	3m
Your Best Entry ↑						
Your submission scored 0.72248, which is not an improvement of your best score. Keep trying!						

0.77990

Keras

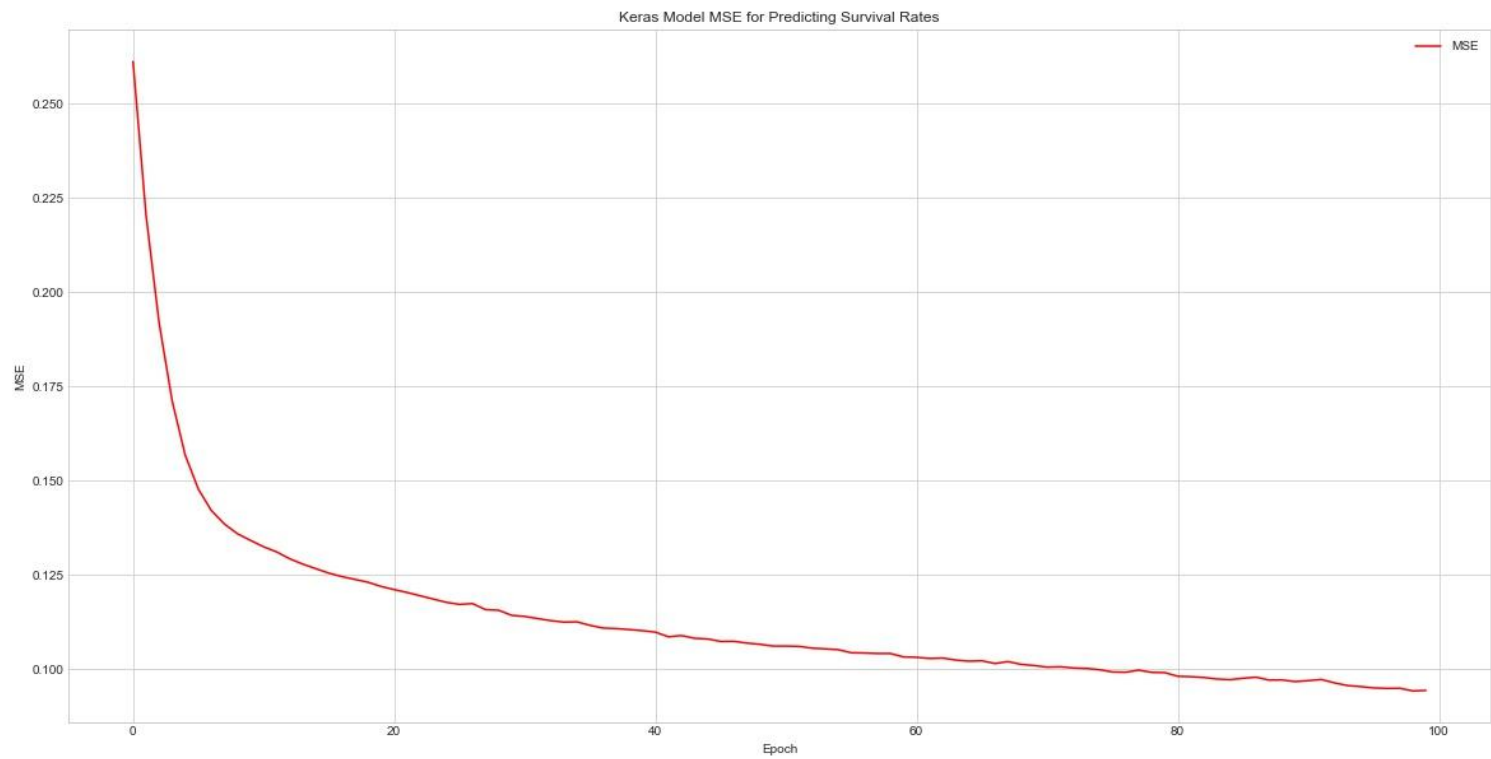
```
model.add(Dense(30, input_dim=X_train.shape[1], activation='relu'))
```

```
model.add(Dense(10, activation='relu'))
```

```
model.add(Dense(1, activation='sigmoid'))
```

```
model.compile(optimizer='adam' , loss='binary_crossentropy', metrics=['accuracy','MSE'])
```

At Epoch 100 Accuracy = 0.8754, MSE = 0.0943, Value Loss = 0.498, Value Accuracy = 0.7982, Value MSE = 0.1499



Thank You