

# HOMWORK FIVE PROPOSAL

CMPS242 - FALL 2015

Andrei Ignat

Sabina Tomkins

Guangyu Wang

## Data Exploration

The first thing we wanted to look at was what the relative importance of different features are. We started this by looking at the correlation between different features. Here is what we saw:

	Survived	Pclass	Age	SibSp	Parch	Fare
Survived	1.000000	-0.323533	-0.043385	0.085915	0.133933	0.266229
Pclass	-0.323533	1.000000	-0.286081	-0.039552	-0.021019	-0.573531
Age	-0.043385	-0.286081	1.000000	-0.142746	-0.200112	0.093249
SibSp	0.085915	-0.039552	-0.142746	1.000000	0.425241	0.358262
Parch	0.133933	-0.021019	-0.200112	0.425241	1.000000	0.330360
Fare	0.266229	-0.573531	0.093249	0.358262	0.330360	1.000000

From this table we can see that if we want to be predicting survival, some especially correlated attributes are fare and Pclass. Both of these things have to do with socio-economic status.

This prompts us to look at the percentage of each class which survived, which results in the following:

	Class One	Class Two	Class Three
Average Survival	.63	.47	.24