# Linear Regression Consulting Project



Now let's set you loose on your first consulting project!

• You've been contracted by Hyundai Heavy Industries to help them build a predictive model for some ships.

You've been flown out to their HQ in Ulsan, South Korea!

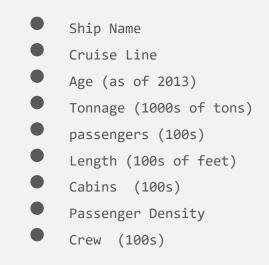


It's one of the world's largest manufacturers of large ships, including cruise liners!



- They need your help them give accurate estimates of how many crew members a ship will require.
- They are currently selling ships to some new customers and want you to create a model and use it to predict how many crew members the ships will need.

They provided you data with these features:



- Your job is to create a regression model that will help predict how many crew members will be needed for future ships.
- In other words, use the features you think will be useful to predict the value in the Crew column.

The client also mentioned that they have found that particular cruise lines will differ in acceptable crew counts, so it is most likely an important feature to include in your analysis!

- The cruise line value is a string however!
- We haven't covered exactly how to convert strings to numbers with Python and Spark (yet)
- Try to see if you can discover how to use **StringIndexer** from the documentation!

- As in any real world project, there are no "100% correct" answers here.
- Just try your best to build the model!
- You can optionally see if you can figure out **StringIndexer** on your own (we'll cover it more formally in future lectures)

- Best of luck with the project, all of the necessary information for the project can be found in the files:
- Linear\_Regression\_Consulting\_Project.ipynb
- cruise\_line\_info.csv