

问题：有一个船厂在造船，希望你帮忙创建一个模型，这个模型能够根据船的一些情况预测这艘船里应该放多少船员。

下面是有关数据集的情况

该数据集包含了158只大型邮轮的尺寸，容积，船员个数及其船员年龄的信息。

自变量：

Ship Name

Cruise Line

Age (as of 2013)

Tonnage (1000s of tons)

passengers (100s)

Length (100s of feet)

Cabins (100s)

Passenger Density

因变量：

Crew (100s)

数据集存放在“cruise_ship_info.csv”文件中. 请利用pySpark建立一个线性回归模型预测船只需要多少船员。