

Proyecto de Consultoría: Regresión Lineal con Python y Spark

Luis R. Furlán

Octubre de 2025

Python y Spark

Ahora quedan libres para que trabajen en su primer proyecto de consultoría!

Han sido contratados por Hyundai Heavy Industries para ayudarles a desarrollar un modelo predictivo para algunos de sus barcos.

Python y Spark

Han sido llevados
a la casa matriz
en Ulsan, Corea
del Sur!



Python y Spark

Es uno de los fabricantes más grandes del mundo de barcos grandes , incluyendo cruceros de línea!



Python y Spark

Quieren que les ayuden a dar estimaciones precisas sobre cuántos miembros de tripulación requerirá un barco.

Actualmente están vendiendo barcos a algunos clientes nuevos y quieren que ustedes desarrollen un modelo y lo utilicen para predecir el número necesario de miembros de la tripulación.

Python y Spark

Les han facilitado las siguientes características:

- Nombre del barco
- Línea de cruceros
- Antigüedad (al 2013)
- Tonelaje (1000s de toneladas)
- pasajeros (100s)
- Longitud (100s pies)
- Cabinas (100s)
- Densidad de pasajeros
- Tripulación (100s)

Python y Spark

Los datos están en el archivo: `barcos_crucero.csv`

Descripción: Medidas del tamaño del barco, capacidad, tripulación, y antigüedad para 158 barcos crucero.

Variables/Columns

Nombre del barco 1-20

Línea de Cruceros 21-40

Antigüedad (al 2013) 46-48

Tonelaje (1000s de toneladas) 50-56

pasajeros (100s) 58-64

Variables/Columns

Longitud (100s de pies) 66-72

Cabinas (100s) 74-80

Densidad de Pasajeros 82-88

Tripulación (100s) 90-96

Python y Spark

Su tarea es crear un modelo de regresión lineal que ayude a predecir cuantos miembros de tripulación serán necesarios para futuros barcos.

Python y Spark

El cliente también mencionó que han encontrado que algunas líneas particulares de cruceros difieren en el conteo aceptable de los miembros de la tripulación. Es posible que esto sea una característica importante a incluir en su análisis!

Python y Spark

Como en cualquier proyecto del mundo real,
no existen respuestas 100% correctas.

Solo intenten desarrollar el mejor modelo.

Python y Spark

Este es un proyecto bastante abierto.

Puede haber más de una forma de crear el modelo!