



COST PREDICTION MODEL

A linear regression machine learning model for estimating medical costs

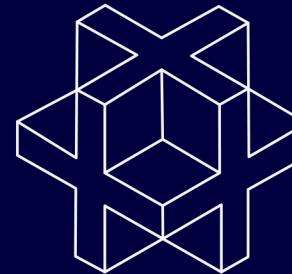


PRESENTED TO
Python Showcase

PRESENTED BY
Will Schultz

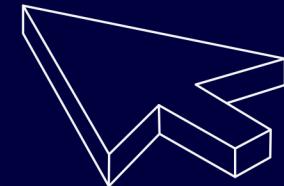
December, 2023

Objectives



MACHINE LEARNING

Apply ML concepts to
HealthPartners data



PYTHON & PYSPARK

Acquire knowledge of libraries,
visualizations and notebooks



A CONCEPTUALIZED APPROACH

Ideas for dataframe analysis + steps for
machine learning

Agenda

“Learning is the only thing the mind never exhausts, never fears, and never regrets.”

- Leonardo da Vinci



Overview of Jupyter notebook:

1

- Python libraries
- Dataframe analysis
- Visualizations
- Checklist for linear regression
- Model predictions (Demo)

2 Questions + Discussion

3

Databricks notebook run through:

- Data overview
- Viewing a dataframe
- Dataframe analysis *Display function
- Preparing the data
- Determine column correlation *Visualization
- Steps to build the ML model

4 Closing Comments + Share Resources

Resources

Learning resources for
Python, PySpark & Databricks

1 Python libraries:

- scikit-learn (<https://scikit-learn.org/stable/>)
- seaborn (<https://seaborn.pydata.org/>)
- matplotlib (<https://matplotlib.org/>)
- pandas (<https://pandas.pydata.org/>)
- jupyter notebook (<https://jupyter.org/>)

2 PySpark documentation:

- pyspark.ml package (<https://spark.apache.org/docs/latest/api/python/reference/pyspark.ml.html>)
- MLlib (<https://spark.apache.org/docs/latest/api/python/reference/pyspark.ml.html>)

3 Databricks documentation:

- Machine Learning (<https://docs.databricks.com/en/machine-learning/index.html>)