

Name: **[Ejjipurapu Rajasekhar](https://learning.edx.org/course/course-v1:IBM+DS0720EN+2T2021/block-v1:IBM+DS0720EN+2T2021+type@sequential+block@cbb408fec2b548e28c1989ff47f4edd8/block-v1:IBM+DS0720EN+2T2021+type@vertical+block@cea0c2ae77c6453ab3c838b2b049fc93)**

Date:24-05-2024

GITHUB PROFILE:https://github.com/rajasekhar0911 4

# Outline

* ExecutiveSummary
* Introduction
* Methodology
* Results
* Visualization–Charts
  + Discussion
* Findings&Implications

•Conclusion•Appendix

2

# ExecutiveSummary

1. DataCollection&Preparation: Utilized public SpaceX API and Wikipedia page.Created'class'columnforsuccessfullanding classification. Explored data usingSQL,visualization,Foliummaps,and dashboards. Selectedrelevantfeaturesformachinelearning.
2. 2. Data Preprocessing:Applied onehot encoding to categorical variables.Standardized data for uniform scale.Optimized model parameters usingGridSearchCV.
3. 3.Machine Learning Models: Developedmodels: LogisticRegression SupportVectorMachine DecisionTreeClassifier KNearestNeighbors Achievedconsistentaccuracy(~83.33%).
4. 4.Evaluation&Analysis: Modelstendedtooverpredictsuccessfullandings.

Identifiedneedformoredatatoenhanceaccuracy.

1. 5.Model PerformanceVisualization: Visualizedaccuracyscorestocompare

modelperformance.

3

# Introduction

* Background:
* Commercialspaceageisbooming.
* SpaceXofferscompetitivepricing($62Mvs.$165MUSD)duetorocket

recovery.

*  Space Y aims to rival SpaceX.Problem:Stage1recovery.Approach:

SpaceYseeksamachinelearningmodeltopredictsuccessfulDatacollectionfromSpaceXAPIandindustrysources.

Preprocessdataandengineerfeatures.

TrainMLmodels:logisticregression,SVM,decisiontrees

4

# Methodology

* 1. DataCollection:

CombineddatafromSpaceXAPI andWikipedia.

* 1. DataWrangling:

Cleanedandorganizedcollecteddata.

* 1. Classification:

Identifiedsuccessfulandunsuccessfullandings.

* 1. ExploratoryDataAnalysis(EDA):

UsedvisualizationandSQLforinsights.

Visualizeddatadistribution.

ExtractedinsightswithSQL.

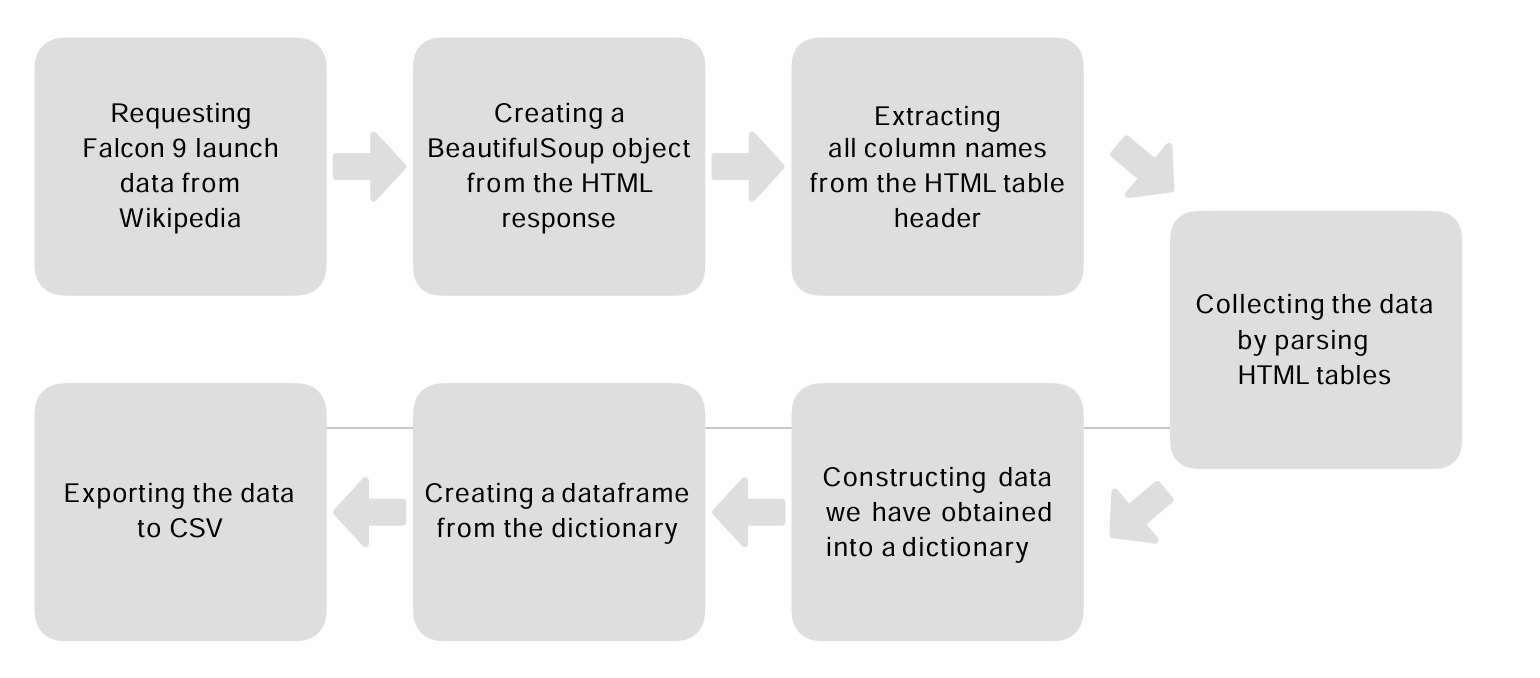
* 1. InteractiveVisualAnalytics: 5

EmployedFoliumandPlotlyDash.

# DataCollection

6

# DataCollection–SpaceXAPI



7

# EDAwith DataVisualization

* EDAwithvisualizationoffersinsightsintodatacharacteristics,aidingindecision-makingandhypothesisgeneration.
* Visualizationshelpidentifypatterns,trends,outliers,anddependencies,enhancingdataunderstanding.
* Findingsguidesubsequentanalysisandmodeling,interpretabilityandrobustnessofresults.
* [GitHubLink:https:/edx/jupyter-labs-EDA-with-Visualization-lab-EvanjaliYaddanapudi.ipynb](https://d.docs.live.net/a7186ef5fc9a3acf/Desktop/edx/jupyter-labs-EDA-with-Visualization-lab-venkatasashank.ipynb)

8

# EDAwithSQL

* Utilized SQL queries to perform comprehensive exploratorydataanalysis(EDA),extractingvaluableinsightsdirectlyfromthedataset.
* SQL facilitated efficient querying, aggregation, andmanipulation of data, enabling in-depth analysis of variousaspectssuchasdistribution,relationships,trends,andoutliers.
* The EDAwith SQL provided a solid foundation for understandingthedataset'scharacteristicsandinformingsubsequentanalyticaldecisions.
* [GitHubLink:-https:/edx/jupyter-labs-eda-sql-edx\_sqllite-EvanjaliYaddanapudi.ipynb](https://d.docs.live.net/a7186ef5fc9a3acf/Desktop/edx/jupyter-labs-eda-sql-edx_sqllite-venkatasashank.ipynb) 9

# BuildanInteractiveMapwithFolium

* + UtilizedFolium,aPython library forcreating interactive maps,toperformgeospatialanalysis andvisualization ofdata.Popup information windowswereincorporated to displayadditionaldetailswhen usersinteracted withmapmarkers, enhancing dataexploration. Interactive features such aszooming,panning, and togglinglayers were integrated toprovide userswithadynamicand
  + GitHubFindings:
  + MapGeneration
  + MarkerClustering
  + PopupInformation
  + [GithubLink:-https:/edx/jupyter-labs-Interactive-Visual-Analytics-with-Folium-lab-EvanjaliYaddanapudi.ipynb](https://d.docs.live.net/a7186ef5fc9a3acf/Desktop/edx/jupyter-labs-Interactive-Visual-Analytics-with-Folium-lab-venkatasashank.ipynb)

10

# BuildaDashboardwithPlotlyDash

* TheInteractiveDashboardbuiltwithPlotlyDashoffersadynamicanduser-friendly

interfaceforexploringandvisualizingdata.

* DataVisualization:
* ImplementedinteractivechartsandgraphsusingPlotlytovisualizekeyinsightsand

trends.UserInteraction:

* Includedlinecharts,barcharts,scatterplots,andheatmapstorepresentdifferent

aspectsofthedata.

* Integrateddropdownmenus,sliders,anddatepickerstoenableuserstofilterand

customizethedisplayeddatadynamically.

* GitHubLink:-[https:/PlotyDash-EvanjaliYaddanapudi.ipynb](https://d.docs.live.net/a7186ef5fc9a3acf/Desktop/edx/Hands-on%20Lab_%20Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash-venkatasashank.ipynb)

11

# PredictiveAnalysis(Classification)

* TheMachineLearningPredictionLabisdedicatedtodevelopingand evaluating predictive models using advanced machinelearningtechniques.
* ModelEvaluation:
* Employedcross-validationtechniquestoassessmodelgeneralizationandrobustness.
* Identifiedkeyfactorsinfluencingthetargetvariablebasedonfeatureimportanceanalysis.
* GitHubLink:-[https:/edx/jupyterlite-EvanjalliYaddanapudi.ipynb](https://d.docs.live.net/a7186ef5fc9a3acf/Desktop/edx/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite-venkatasashank.ipynb)

12

# Results

13

# FlightNumbervs.LaunchSite

14

# AllLaunchSiteNames

* + Findings
  + Finding1:Relational databasessuchas MySQLandPostgreSQLcontinue tobe widely adopted for traditional datamanagementtasksduetotheirrobustnessandstability.
  + Finding2:NoSQLdatabaseslikeMongoDBandRedis aregainingpopularityforhandlingunstructuredandsemi-structureddata,suchassocialmediaanalyticsandIoTapplications.
  + •Embracecloud-nativedatabasesandmanagedservicestoleveragethebenefitsofscalability,flexibility,andreducedmaintenanceoverhead,enablingfastertime-to-marketandcost savings.
  + Implications
  + Organizationsshouldmaintainproficiencyinrelationaldatabasestomanagestructureddataeffectively,particularlyfor

legacysystemsandtraditionalapplications.

* + ConsideradoptingNoSQLdatabasesforprojectswithrequirementsforhandlingdiverseandrapidlychangingdatatypes,

suchassocialmediaanalyticsandIoTapplications..

15

# Conclusions

* User-friendlyinterfaceandintuitivedesignenableeasycreationandcustomization

ofdashboards,reducingthelearningcurveforusers.

•Seamlessdataintegrationcapabilitiesensureaccesstocomprehensivedatafromdiversesources,enhancingdataanalysisanddecision-making.

* Interactivevisualizationfeaturesempowerusers

toexploredatadynamically,uncoveringinsightsandtrendsthatdrivebusinessoutcomes.

•Robustcollaborationandsharingfunctionalitiesfacilitateteamworkandcommunication, fostering a data-driven culture within theorganizationanddrivingcollectiveintelligenc

16

