Winning Space Race with Data Science

Name:Veera Prasad Nayak
Date:24-05-2024

https://github.com/Prasad554

Outline

- ExecutiveSummary
- Introduction
- Methodology
- Results
- Visualization-Charts
- Discussion
- Findings&Implications
- Conclusion•Appendix

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.
5. 5.Model PerformanceVisualization: Visualizedaccuracyscorestocompare modelperformance.

Introduction

- Background:
- Commercialspaceageisbooming.
- SpaceXofferscompetitivepricing($62Mvs.$165MUSD)duetorocket recovery.
- Space Y aims to rival

SpaceX.Problem:Stage1recovery.Appro

ach:

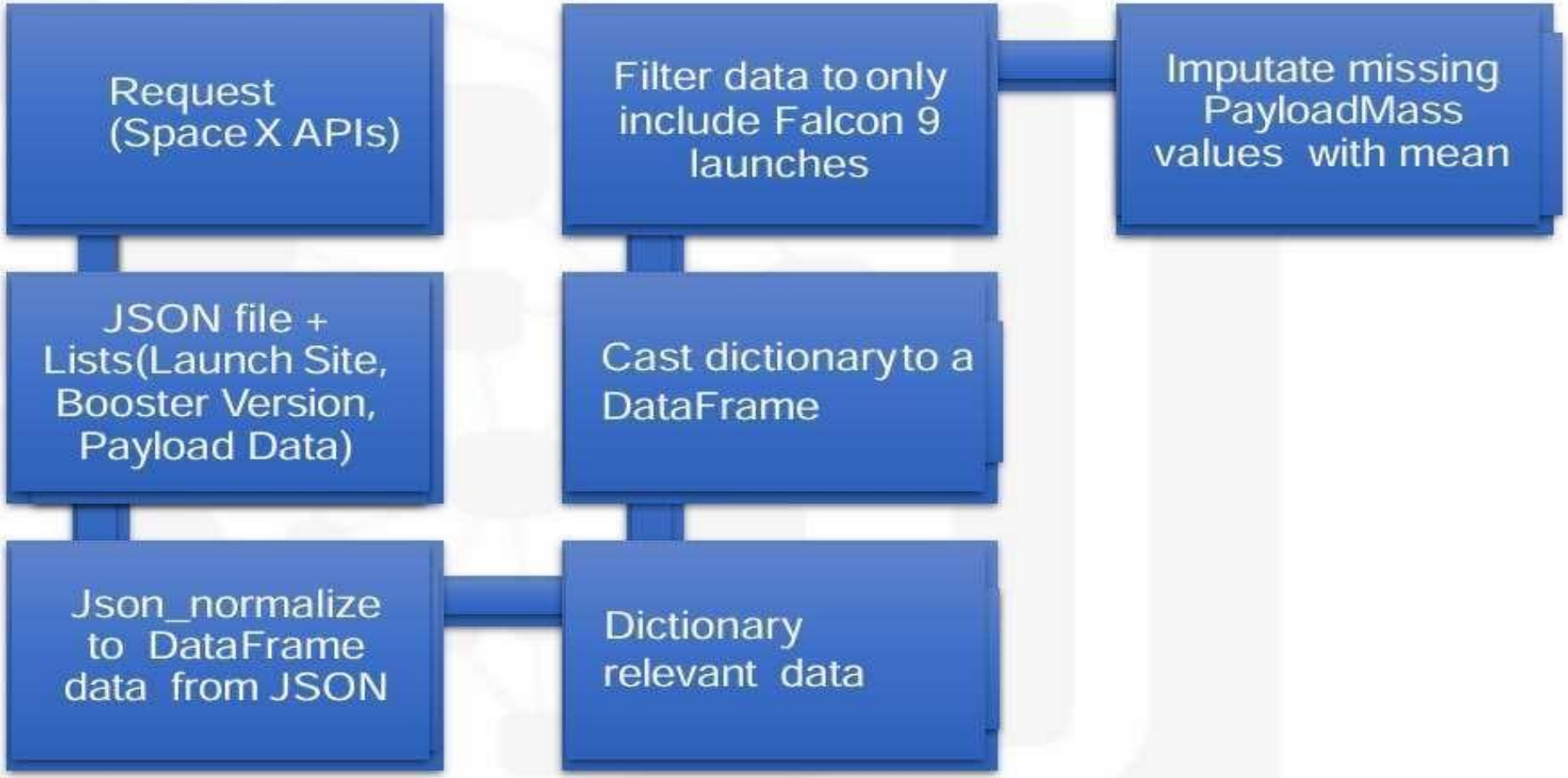
- SpaceYseeksamachinelearningmodeltopredictsuccessfulDatacollecti
- onfromSpaceXAPIandindustrysources.
- Preprocessdataandengineerfeatures.
- TrainMLmodels:logisticregression,SVM,decisiontrees

Methodology

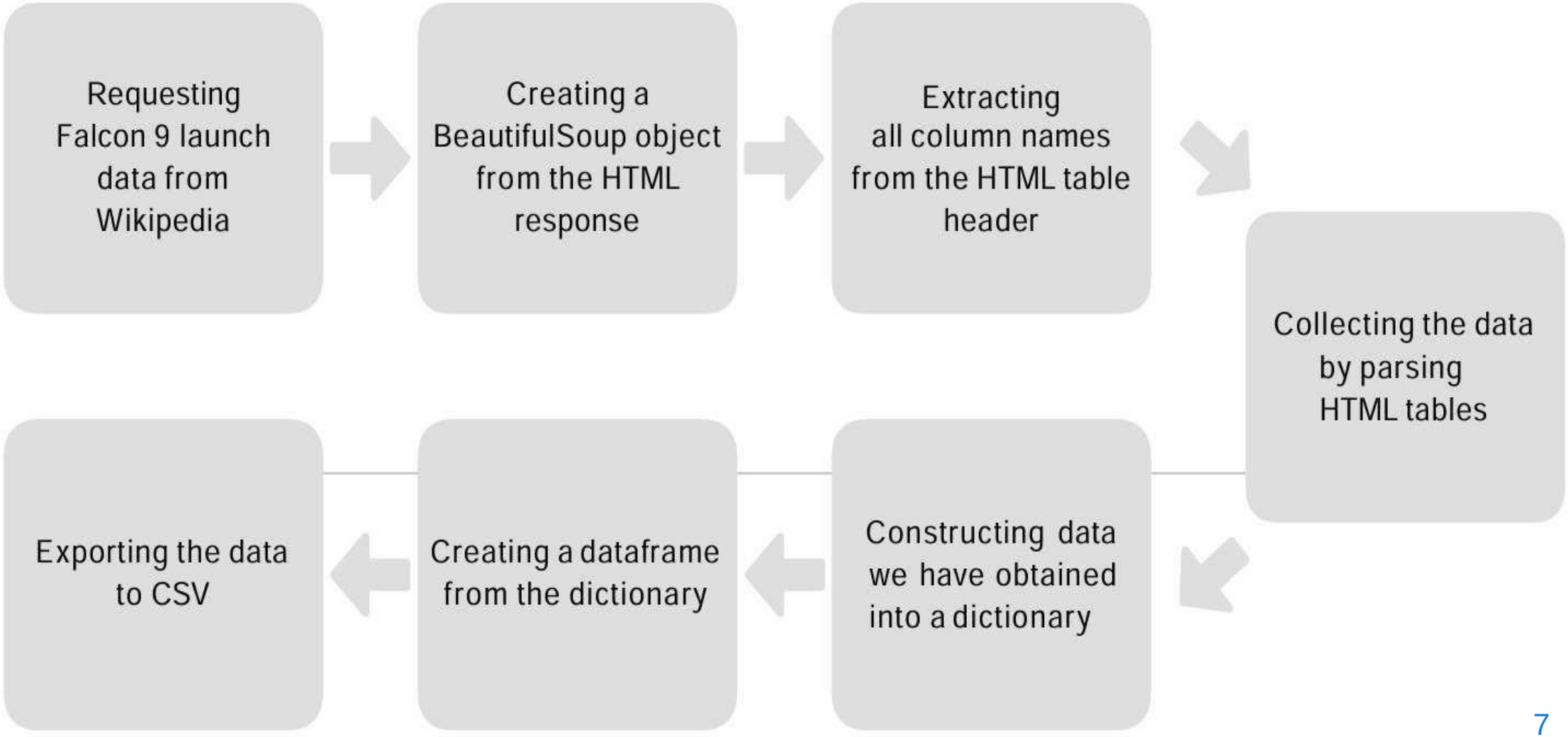
1. DataCollection:
   - CombineddatafromSpaceXAPI andWikipedia.
2. DataWrangling:
   - Cleanedandorganizedcollecteddata.
3. Classification:
   - Identifiedsuccessfulandunsuccessfullandings.
4. ExploratoryDataAnalysis(EDA):
   - UsedvisualizationandSQLforinsights.
   - Visualizeddatadistribution.
   - ExtractedinsightswithSQL.
5. InteractiveVisualAnalytics:
   - EmployedFoliumandPlotlyDash.

DataCollection

RESULTS



DataCollection–SpaceXAPI



EDAwith DataVisualization

- EDAwithvisualizationoffersinsightsintodatacharacteristics,aid ingindecision-makingandhypothesisgeneration.
- Visualizationshelpidentifypatterns,trends,outliers,andde pendencies,enhancingdataunderstanding.
- Findingsguidesubsequentanalysisandmodeling,in terpretabilityandrobustnessofresults.
- GitHubLink:https:/edx/jupyter-labs-EDA-with-Visualization-lab-EvanjaliYaddanapudi.ipynb

EDAwithSQL

- Utilized SQL queries to perform comprehensive exploratorydataanalysis(EDA),extractingvaluableinsightsdirectly fromthedataset.
- 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

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

BuildaDashboardwithPlotlyDash

- TheInteractiveDashboardbuiltwithPlotlyDashoffersadynamicanduser-friendly interfaceforexploringandvisualizingdata.
- DataVisualization:
- Implementedinteractivechartsandgraphsusing Plotlytovisualizekeyinsightsand trends.UserInteraction:
- Includedlinecharts,barcharts,scatterplots,andheatmapstorepresentdifferent aspectsofthedata.
- Integrateddropdownmenus,sliders,anddatepickerstoenableuserstofilterand customizethedisplayeddatadynamically.
- GitHubLink:-https:/PlotyDash-EvanjaliYaddanapudi.ipynb

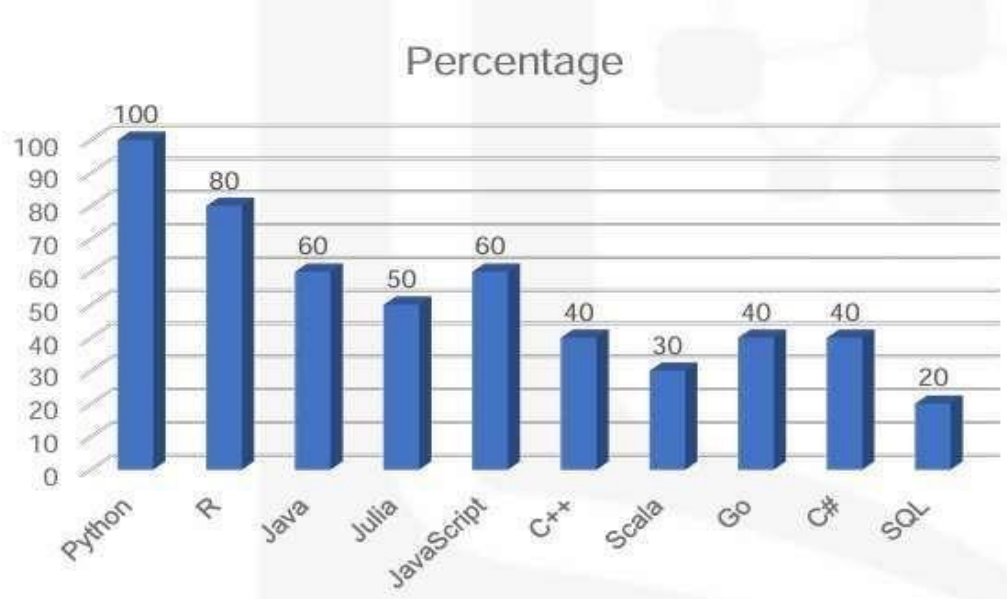
PredictiveAnalysis(Classification)

- TheMachineLearningPredictionLabisdedicatedtodevelopingand evaluating predictive models using advanced machinelearningtechniques.
- ModelEvaluation:
- Employedcross-validationtechniquestoassessmodelgeneralizationandr obustness.
- Identifiedkeyfactorsinfluencingthetargetvariablebasedonfea tureimportanceanalysis.
- GitHubLink:-https:/edx/jupyterlite-EvanjalliYaddanapudi.ipynb

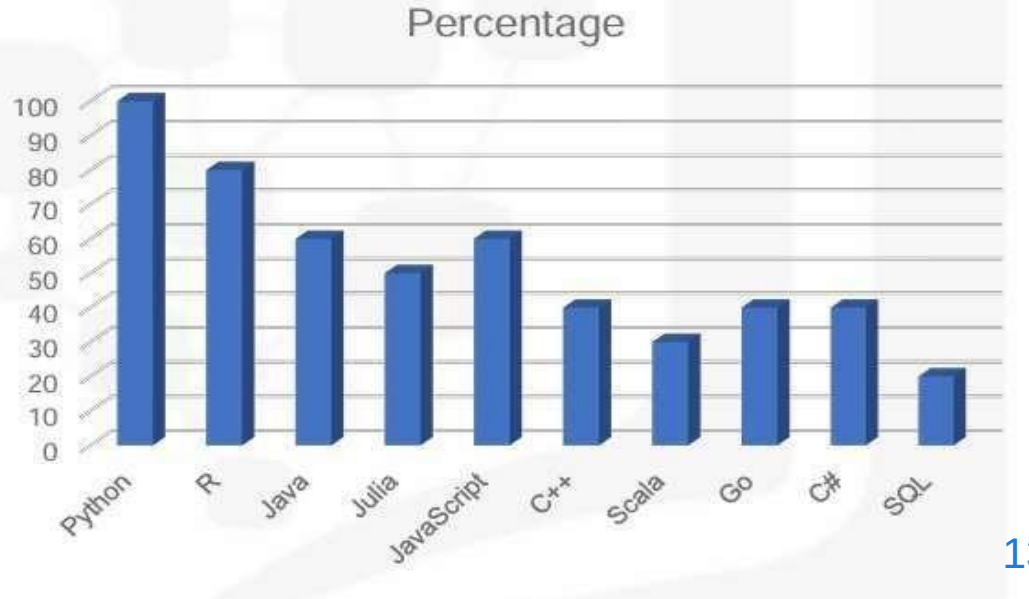
Results

PROGRAMMING LANGUAGE TRENDS

2024



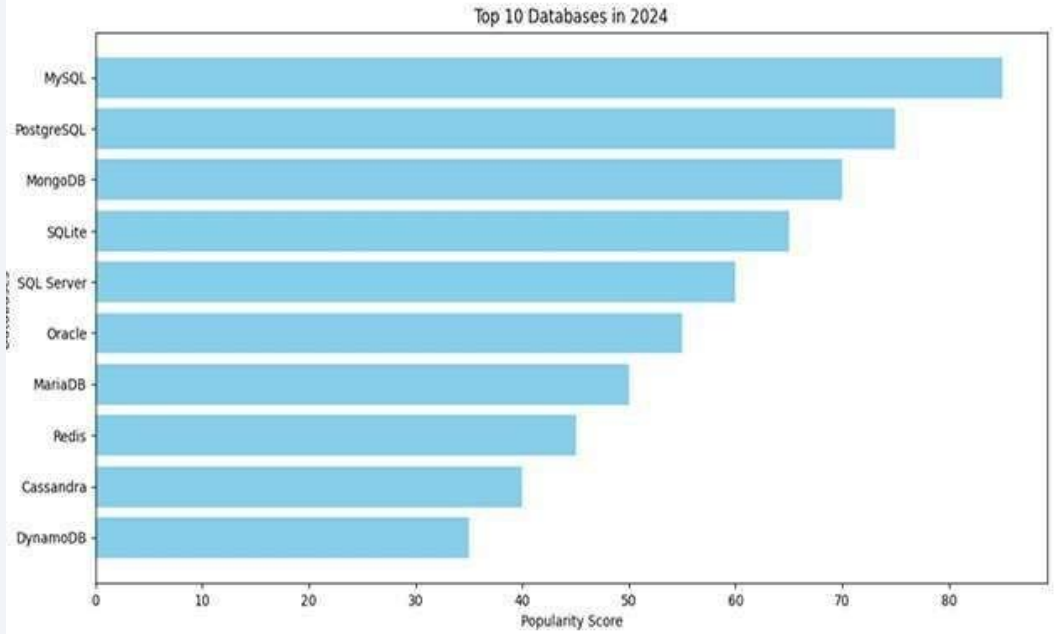
2025



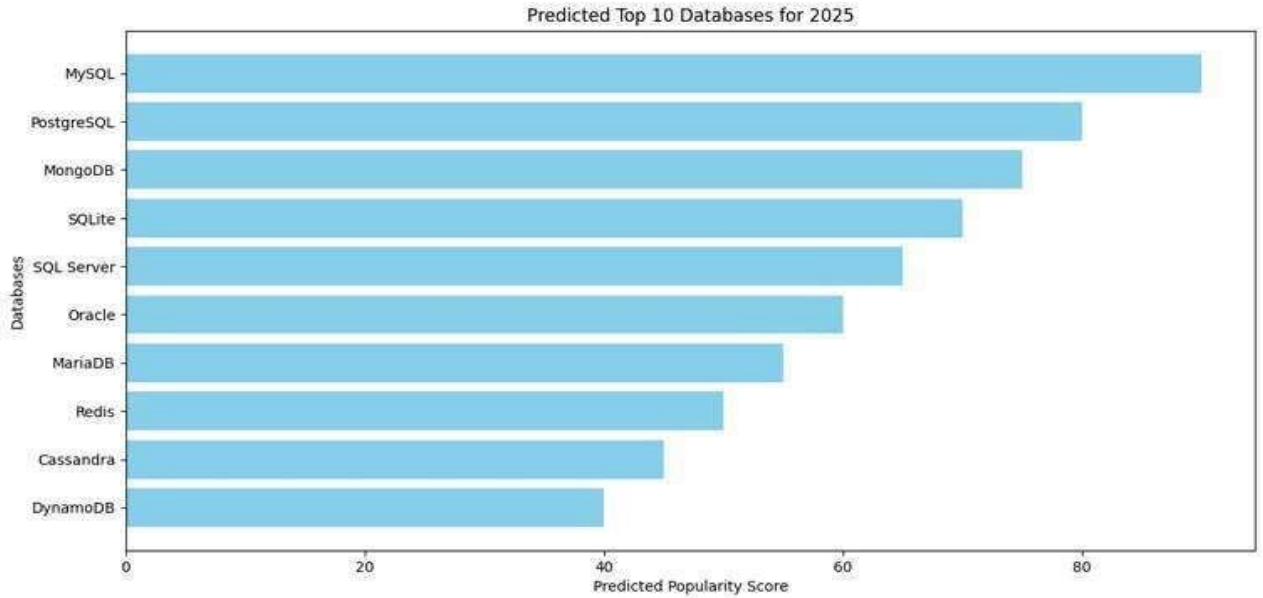
FlightNumbervs.LaunchSite

DATABASE TRENDS

Current Year 2024



Next Year 2025



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..

Conclusions

- User-friendlyinterfaceandintuitivedesignenableeasycreationandcustomization ofdashboards,reducingthelearningcurveforusers.

- Seamlessdataintegrationcapabilitiesensureaccesstocomprehensivedatafromdiverses ources,enhancingdataanalysisanddecision-making.

- Interactivevisualizationfeaturesempowerusers toexploredatadynamically,uncoveringinsightsandtrendsthatdrivebusinessoutcomes.

- Robustcollaborationandsharingfunctionalitiesfacilitateteamworkandcommunication, fostering a data-driven culture within theorganizationanddrivingcollectiveintelligenc

Thank you!