PublictransportationEfficiencyanalysis

IBMCOGNOSVISUALIZATION:

Inthispartyouwillcontinuebuildingyourproject.
ContinuebuildingtheanalysisbycreatingvisualizationsusingIBMCognosandintegratingcodefordataanalysis.
Designdashboardsandreports inIBMCognos tovisualizeon-timeperformance,passenger feedback,andserviceefficiencymetrics. Use code (e.g., Python) to perform advanced data analysis, such as calculating service punctuality rates or sentiment analysis on passenger feedback.

TEAMMATES:

S.Surya(952321205701) https://github.com/Suryas9503/DAC_Phase5.git
G.Ramkumar(952321205305) https://github.com/Ramkumarking/Ramkumar.git
M.Akash(952321205702) https://github.com/Akashman0/Akashmano.git
N.Manikandan(952321205302) https://github.com/Manik1308/Manikandan.git

ContinuebuildingtheanalysisbycreatingvisualizationsusingIBMCognosandintegratingcodefordataanalysis:

1. DataAnalysiswithCode:

Python Example:
Assumingyouhave yourdataina format likeCSVoradatabase,youcanusePythonwithlibrariessuchasPandas,Matplotlib, Seaborn,orPlotlyfordataanalysisandvisualization.

Jupyternotebook:

ConsiderusingJupyterNotebooks for interactiveanalysisandvisualization.Itallowsyoutocombinecode,visualizations,and explanatory text.

```
importpandasas pd
```

```
importmatplotlib.pyplotasplt

#Loadyourdata
data=pd.read_csv('your_data.csv')

# Perform some basic analysis
summary_stats=data.describe()

#Createahistogram
plt.hist(data['your_column'],bins=20,color='blue',alpha=0.7)
plt.title('Distribution of Your Column')
plt.xlabel('Values')
plt.ylabel('Frequency')
plt.show()
```

Dataset:

TripID	RouteID	StopID	StopName	WeekBeginning	Numberofboardings
23631	100	14156	181	CrossRd30-06-2013	00:00
23631	100	14144	177	CrossRd30-06-2013	00:00
23632	100	14132	175	CrossRd30-06-2013	00:00
23633	100	12266	ZoneA	ArndaleInterchange	30-06-201300:00
23633	100	14147	178	CrossRd30-06-2013	00:00

2. VisualizationwithIBMCognos:

IBMCognos Analytics:

ConnecttoDataSource:

ImportyourdataintoIBMCognos Analytics.

Definerelationshipsbetweentablesifapplicable. Create
Reports and Dashboards:

Usethedrag-and-dropinterfacetocreatereports.
Incorporatevariousvisualizationslikecharts,graphs,andtables.
Create dashboards to bring multiple visualizations together.
InteractiveFeatures:

Leverageinteractivefeatures inCognos,suchasdrill-throughoptions fordetailedanalysis. Customization:

Customizetheappearanceofyourvisualizationstoalignwithyouranalysisneeds.

3. CombineCodeandCognos:
Ifyouwanttointegratecode-basedanalysisresultsintoCognos:

ExportAnalysisResults:

Afterrunning yourPython/Ranalysis, exporttheresultstoaformatcompatiblewithCognos(e.g.,CSV). Import into Cognos:

ImporttheresultsintoCognosasanewdataset. Create
Visualizations:

BuildvisualizationsinCognosusingtheimporteddataalongsideotherdatasources.

4. ScheduleandAutomate:
Considerautomatingtheprocess:

ScriptAutomation:

AutomateyourPython/Rscriptsusingtools likecronjobsortaskscheduler.

CognosAutomation:

ScheduledatarefreshesandreportgenerationinCognos. Important

Note:

Ensurethat yourIBMCognosenvironment isproperlyconfiguredtoworkwithexternaldatasourcesandcode integration.Detailed steps may vary based on the versions and configurations of the software you are using.

Explainhowtheinsightsfromtheanalysiscansupporttransportationimprovementinitiatives.
Analyzingdataand gaining insightscansignificantlycontributetotransportationimprovement initiativesinvariousways.Here'sa breakdown of how insights from analysis can support such initiatives:

IdentifyingTrafficPatterns:

Insight:Analyzingtrafficdatahelpsinunderstandingpeaktraffictimes,congestion-proneareas,andtrafficflowpatterns.
Support forImprovement:Thisinformation canbeusedtooptimizetrafficsignaltimings,implement congestionpricing,orplanfor additional lanes or alternative routes in busy areas.

PredictiveMaintenance:

Analyzingmaintenancedataforvehiclesandinfrastructurehelpspredictwhenmaintenanceislikelytobeneeded.
Support forImprovement:Proactive maintenancecanbescheduled,reducingthechancesofbreakdownsor failuresthatcandisrupt transportation systems.
OptimizingRoutes:

Dataanalysisofhistoricaltrafficdatacanrevealthemostefficientroutesfordifferenttypesoftransportation.
Support forImprovement:Thisinformation canbeusedtooptimizepublictransportationroutes,planfornew infrastructure, and enhance overall transportation efficiency.

DemandForecasting:

Analyzinghistoricalanddemographicdatahelpsinpredictingfuturetransportationdemands.
Support forImprovement:Transportationagenciescanusethisinformationtoplanfor future infrastructureneeds, allocateresources efficiently, and design systems that can adapt to changing demand patterns.

PublicTransportPlanning:

Insight:Studyingpublictransportusagepatternscanprovideinsightsintotheeffectivenessofexistingsystems.
Support forImprovement:Thisinformation canguidethedevelopmentofnewroutes, adjustmentofschedules, and improvement of overall public transportation services.
EnvironmentalImpactAssessment:

Insight:Analyzingtransportationdatacanhelpassesstheenvironmentalimpactofdifferentmodesoftransport.
Support forImprovement:Thisinformation iscrucialfor developingsustainabletransportationinitiatives,suchaspromotingpublic transport, cycling, or implementing electric vehicles.

safetyImprovements:
Analyzingaccidentdatahelpsidentifyhigh-riskareasandcommoncausesofaccidents.
Support forImprovement:Resourcescanbedirectedtowardsimprovingroadsafetyinspecific locations,implementingtraffic calming measures, and enhancing law enforcement in critical areas.

SmartTrafficManagement:
Real-timeanalysisoftrafficdataenablesdynamictraffic management.
Support forImprovement:Adaptivetrafficsignalcontrol, dynamic laneassignments, andreroutingbasedonreal-timedatacan optimize traffic flow and reduce congestion.

CostOptimization:

Analyzingoperationalcostsandresourceutilizationdatahelpsidentifyareasofinefficiency.
Support forImprovement:Transportationagenciescanstreamlineoperations,reducecosts,andallocateresources moreeffectively based on the insights gained.

InfrastructurePlanning:

Analyzing data on population growth, urban development, and economic trends helps in long-term infrastructure planning.
Support forImprovement:Thisinformation isessentialfordevelopingandexpandingtransportationinfrastructureto meet future demands.

ProjectObjective:
Objective:EnhanceUrbanTransportationEfficiency

Overview:
Theprojectaimsto improve urbantransportationefficiencybyleveragingdata-driveninsightsandtechnology.The focus ison optimizing traffic flow, enhancing public transport services, and implementing predictive maintenance for vehicles.

DesignThinkingProcess:
Empathize:

Understandthechallengesfacedbycommuters,trafficmanagementauthorities,andpublictransportoperators. Conduct surveys, interviews, and observational studies to gather insights.

Define:
Clearlydefinetheprojectgoalsandkeyperformance indicators(KPIs).
Identifypainpointsandchallengesinthecurrenttransportationsystem.

Ideate:
Brainstormpotentialsolutionstoaddressidentifiedchallenges.
Encouragecreativityandcollaborationamongteammemberstogeneratediverseideas.

Prototype:
Developprototypesofproposedsolutions.
Usesimulationtoolstomodeltheimpactofproposedchangesontrafficflow,publictransportefficiency,andmaintenanceprocesses.

Test:

Testprototypesincontrolledenvironmentsorthroughsimulations.
Gatherfeedbackfromstakeholdersandrefineprototypesbasedonthetestingresults.

DevelopmentPhases:
Phase1:DataCollectionandIntegration

Tasks:
Identifyrelevantdatasources(trafficcameras,GPSdevices,maintenancelogs). Develop scripts for real-time data collection and processing.
Integratedataintoacentralizeddatabase.

Phase2:PredictiveMaintenanceModelDevelopment Tasks:
Usehistoricalmaintenancedatatodeveloppredictivemaintenancemodels.
Implement machine learningalgorithms for predicting maintenance needs.
Integratethepredictive maintenancesystemwiththeoverallinfrastructure.

Phase3:TrafficFlowOptimization Tasks:
Analyzehistoricaltrafficdatatoidentifycongestionpatterns. Develop algorithms for optimizing traffic signal timings.
Implementareal-timetrafficflowmonitoringsystem.

Phase4:PublicTransportEnhancement
Tasks:
Analyzepublictransportdatatoidentify inefficiencies.
Optimizepublictransportroutesandschedulesbasedonridershippatterns.
Implementreal-timetrackingandcommunicationsystemsforpublictransportvehicles.

Phase5:IBMCognosIntegrationforDataVisualization Tasks:
CustomizeIBMCognosdashboardsfortrafficflow,publictransportefficiency,andmaintenanceinsights.

IntegrateIBMCognoswiththecentralizeddatabaseforseamlessdatavisualization.

Phase6:TestingandOptimization Tasks:
Conductextensivetestingoftheentiresystem.
Optimizealgorithms,dashboards,andoverallsystemperformancebasedontestingresults.

Phase7:DeploymentandTraining Tasks:
Deploy the system in a controlled environment.
Conducttrainingsessionsforusersandstakeholders.

Phase8:ContinuousImprovement Tasks:
Establishafeedbackloopforcontinuous improvement.
Monitorsystemperformanceandgatheruserfeedbackforfurtherenhancements.

KeyMilestones:
Milestone1:CompletionofDataIntegration(EndofMonth2)
Integrated data sources into a centralized database.

Milestone2:PredictiveMaintenanceModelImplementation(EndofMonth4)
Implemented and tested predictive maintenance models.

Milestone3:Traffic FlowOptimizationSystem(EndofMonth6)
Implemented algorithms for optimizing traffic flow.

Milestone4:PublicTransportEnhancement(EndofMonth8)
Optimized public transport routes and schedules.

Milestone5:IBMCognosIntegration(EndofMonth10) Integrated
IBM Cognos for data visualization.

Milestone6:SystemTestingandOptimization(EndofMonth12)
Conducted thorough testing and optimized the system.

Milestone7:DeploymentandTraining(EndofMonth14)
Deployedthesysteminacontrolledenvironmentandconductedusertraining.

Milestone8:ContinuousImprovementInitiatives(Ongoing)
Established a continuous improvement framework.
SuccessCriteria:

TrafficFlowImprovement:
Percentagereductioninpeak-hourcongestion.

PublicTransportEfficiency:
Increasedridership.
Improvedon-timeperformance.

PredictiveMaintenance:
Reductioninunscheduledvehiclebreakdowns.

UserSatisfaction:
Positivefeedbackfromcommutersandtransportationauthorities.
Byfollowingthisstructuredapproach,theprojectaimstonotonlymeet itsobjectivesbutalsoensure thatthesolutions implemented are user-centric, data-driven, and continuously improved based on real-world feedback and performance metrics.

AnalysisObjectives:
TrafficFlowOptimization:

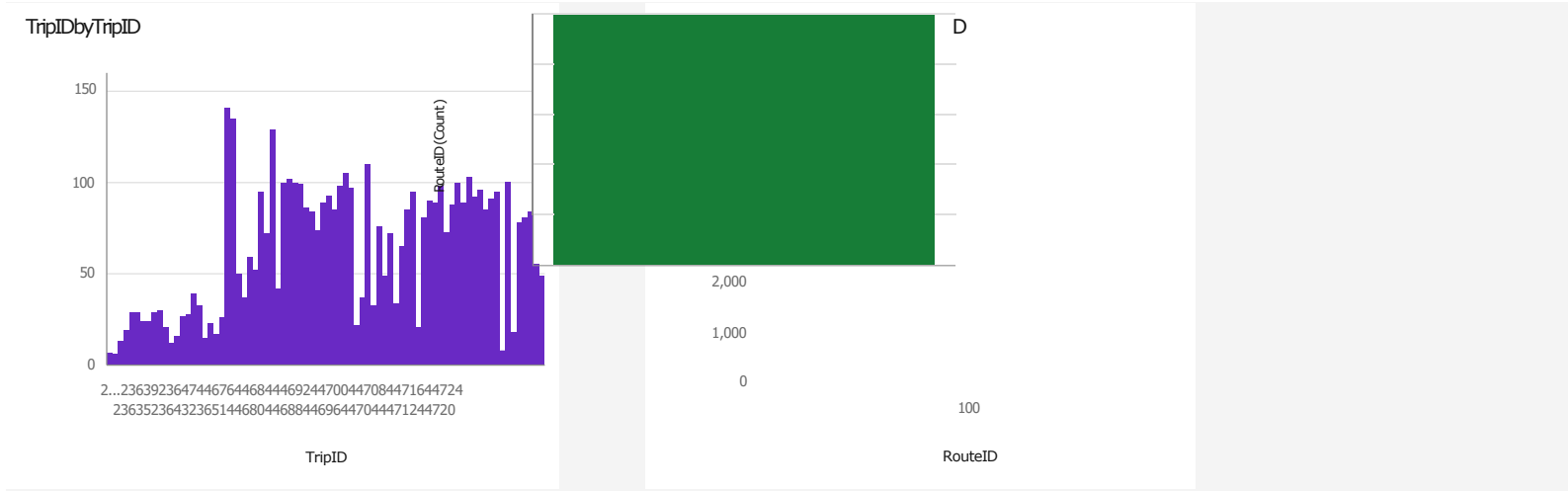
Objective:Improvetrafficflowina citybyidentifyingcongestionpatternsandsuggestingoptimizations. Public Transport Efficiency:

Objective:Enhancepublictransportservicesbyanalyzingridership datatooptimizeroutesandschedules. Predictive Maintenance:

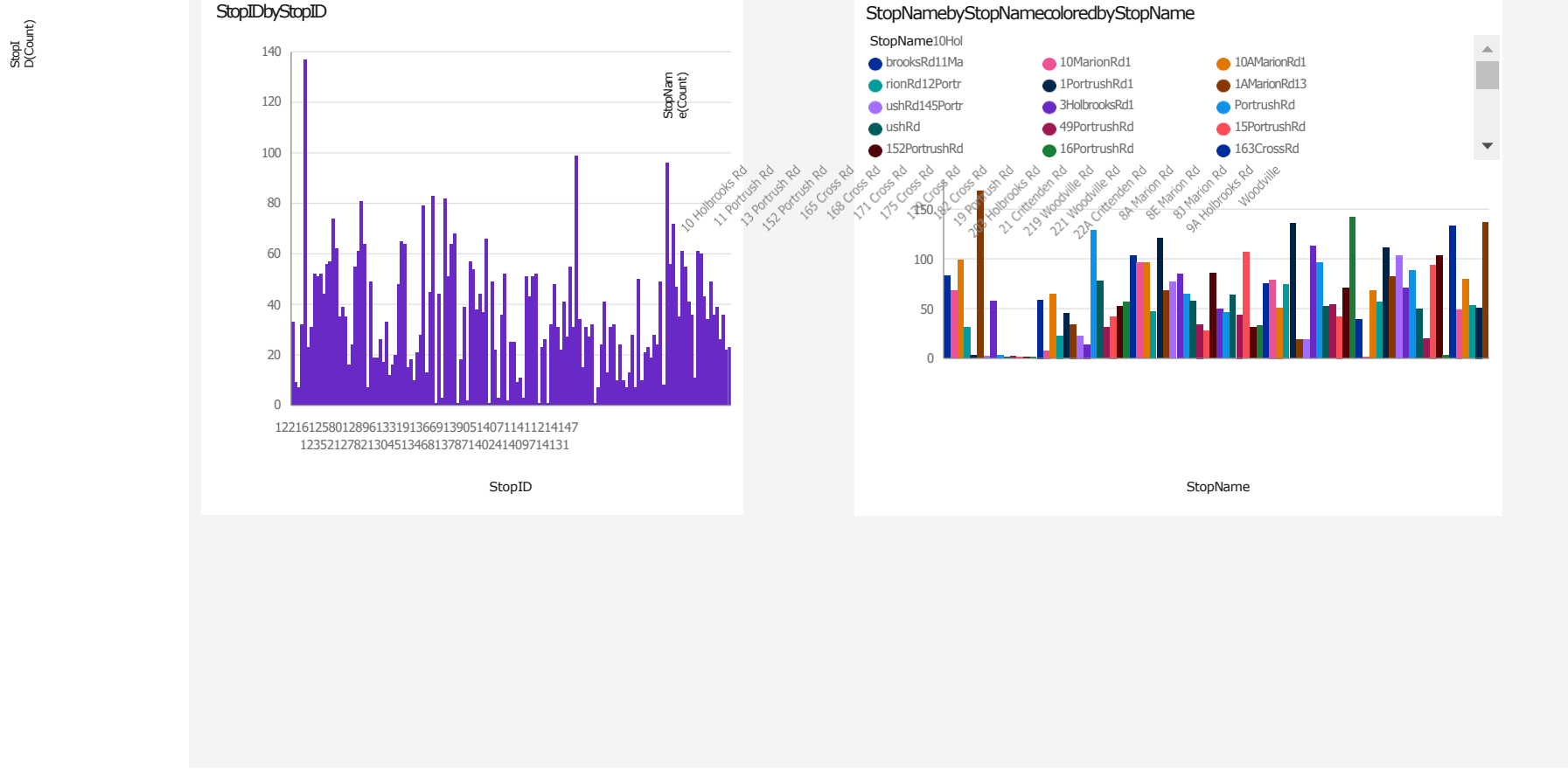
Objective:Reducevehiclebreakdownsbypredictingmaintenanceneedsthroughhistoricalvehicledataanalysis.

Tab1

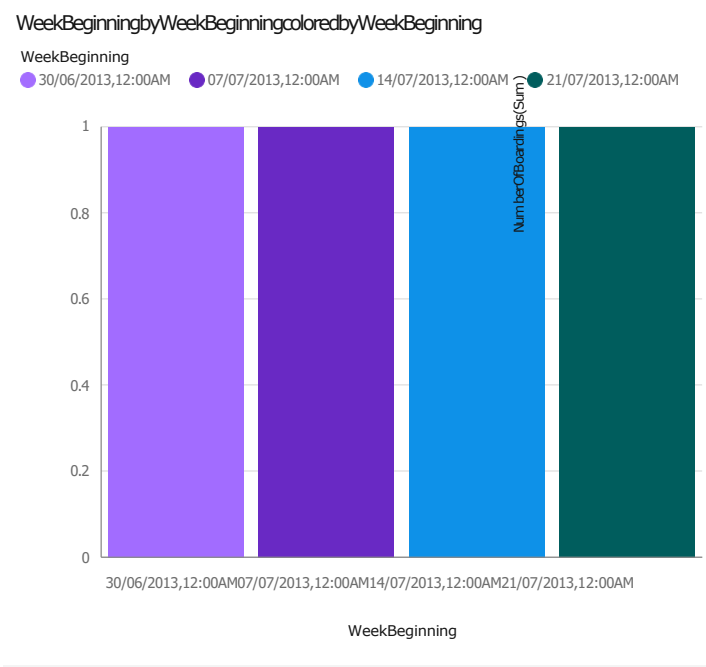
Designdashboardsandreportsin IBM Cognos to visualize on-time performance, passenger feedback, and service efficiency metrics.

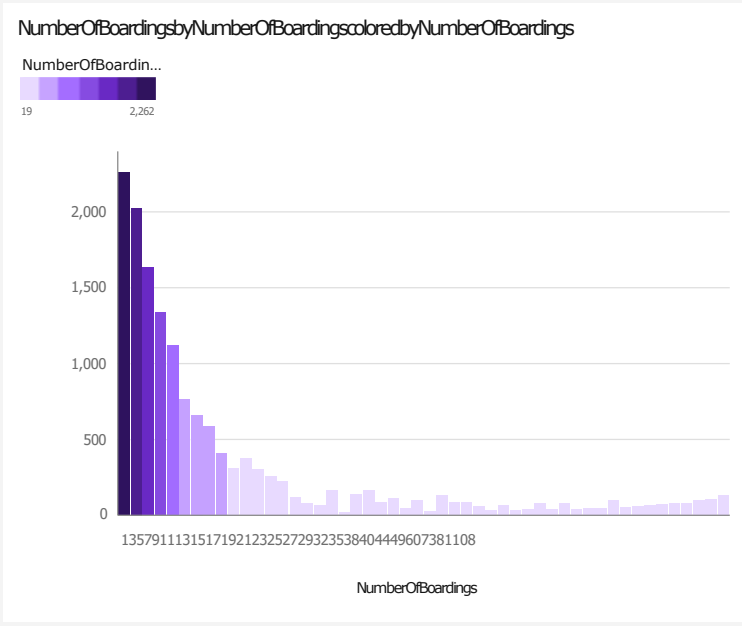


DesigndashboardsandreportsinIBMCognostovisualizeon-timeperformance,passengerfeedback,andserviceefficiencymetrics.



Designdashboardsandreports inIBMCognostovisualizeon-timeperformance,passengerfeedback,andserviceefficiencymetrics.





wherewewilldelvedeeperintotheanalysis,modeling,andrecommendationstoenhancepublictransportationefficiency.Awell-structured dataset can significantly impact the project's overall success.

CalculatingServicePunctualityRates:

DataPreparation:

Ensurethatyourdatasetincludesrelevantinformationlike scheduleddeparturetimes,actualdeparturetimes,andotherrelevant variables.

Cleanandpreprocessthedatatohandlemissingvaluesoroutliers.

CalculatePunctuality:

Createanewvariablethatrepresentsthetimedelay(actualdeparturetime-scheduleddeparturetime).

Analyzethedistributionofdelaysand identifypunctualitythresholds(e.g., departureswithin15 minutesofthescheduledtimeare considered on time).

CalculatePunctualityRate:

Determinetheoverallpunctualityratebydividingthenumberofon-timedeparturesbythetotalnumberofdepartures.

VisualizetheResults:

Createvisualizations(e.g., barcharts, linecharts)torepresentthepunctualityrates overtimeorbyspecificcategories. Sentiment Analysis on Passenger Feedback:

DataCollection:

Collectandcompilepassengerfeedbackdata.Thiscouldbefromsurveys,socialmedia,orothersources.

TextPreprocessing:

Cleanandpreprocessthetextdatabyremovingstopwords,punctuation,andconvertingtexttolowercase. Sentiment Analysis:

Usenaturallanguageprocessing(NLP)libraries likeNLTKorspaCyinPythontoperformsentimentanalysis onthepassenger comments.

Assignsentimentscores(positive, negative,neutral)toeachcomment.

Aggregate Results:

Aggregatesentimentscorestogetanoverallsentimentforeachaspect(e.g.,servicequality,cleanliness).

VisualizeSentimentTrends:

Createvisualizations(e.g.,stackedbarcharts,linecharts) to visualizethesentimenttrendsovertimeorbydifferentservice categories.

ExamplePythonCode(SentimentAnalysis):

```
importpandasaspd
fromnltk.sentimentimportSentimentIntensityAnalyzer data

= pd.read_csv('passenger_feedback.csv')

sid= SentimentIntensityAnalyzer()

def get_sentiment_score(comment):
    sentiment=sid.polarity_scores(comment)
    return'positive'ifsentiment['compound']>=0else'negative'
```

data['sentiment']=data['comments'].apply(get_sentiment_score)

sentiment_counts = data['sentiment'].value_counts()

sentiment_counts.plot(kind='bar',color=['green','red'])
plt.title('Sentiment Analysis of Passenger Feedback')
plt.xlabel('Sentiment')
plt.ylabel('NumberofComments') plt.show()

Remember to adapt the code based on the specific structure of your data and the tools you are using for sentiment analysis. Thesearegeneralguidelines, andtheexactsteps mayvarybasedonthespecifics ofyour datasetandthetools youareusing.

Code:

∅

{x} `import matplotlib.pyplot as plt`
`import seaborn as sns`

j [>] `df = pd.read_csv("/content/20140711.CSV")`

0	2Z031	400	1Ã58.0	181 Cross Rd	2013•06•30 00:00:00	1.0 #
1	?Z£3}	i00	Ñ1440	177 Cross Rd	2013•06•30 00-0000	1.0
2	23632	100	14132.0	175 Cross Rd	2013-06-30 00 00 00	1.0
3	23633	100	12288.0	Zone A Amdale Inierchange	2013-08-30 00.00.00	2.0
¥	23633	Ï00	14147.0	178 Cross Rd	2013-06•30 00:00.00	t0

v `dŁirrfo()`

s	Oolumn	Non-Null Count	Dtype
2	StopID	95801 non-null	float64
S	IJumberOfBoardings	9580t ⌀•n null	Iboa Ł6zt

" [6] `dl. descr*be()`

cour·è	CE802 000000	95801 000000	95804 OOOOD
min	5605.000000	12213.000000	1.000E)(0
M	2536S 500000	13669 000000	2.000000
75ß	44701 000000	14104 000000	4.000000
max	44T41 000000	18072 000000	493.000000

@ `sns.‹ountpiot(¥='3trintJane',data=d-I)`

‹Axes· xIabeI –'JIopNane', yJ abet–'counI'›

000

2500

2000

· Ï300

```
TripID          0
RouteID         1
StopID          1
StopName        1

dtypc: int6#
```

```
    if  df[feature] .isnull(). sum(›e:
        print(J'"(faatura)  :  (round(ay[Taature . lsnutl().aaan( ),4)‘100)¥")
```

```
StopID : 0.0%
StopName : 0.0%
```

[J4] ** find dubticate rows in dataset

```
a«piicate = df[df. avpiicaud()
```

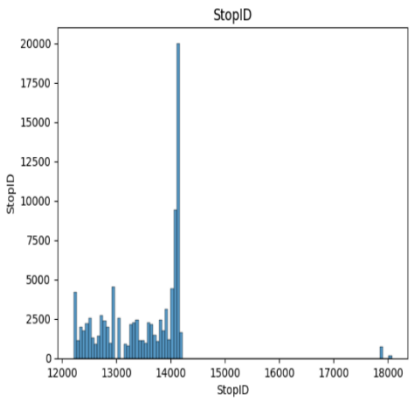
```
TripID  RouteID  StopID  StopName  WeekBeginning  NumberOfBoardings
```

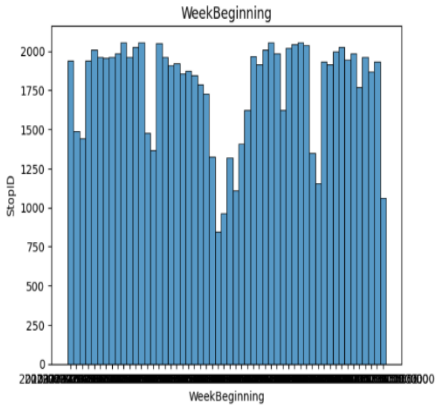
```
print(f" {i}  :  {len(df[i].unique())}")
```

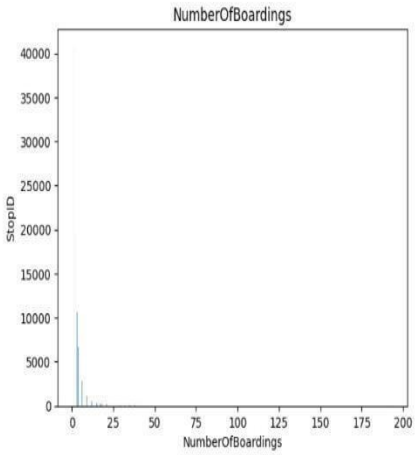
```
TripID  :  182
RouteID  :  7

StopName  :  97

fumbr0%o dines    145
```








```
# removing outliers
Q1 = df.quantile(0.25)
Q3 = df.quantile(0.75)
IQR = Q3 - Q1
```

$ g [27] df[“Stoptlane° J.value counts()

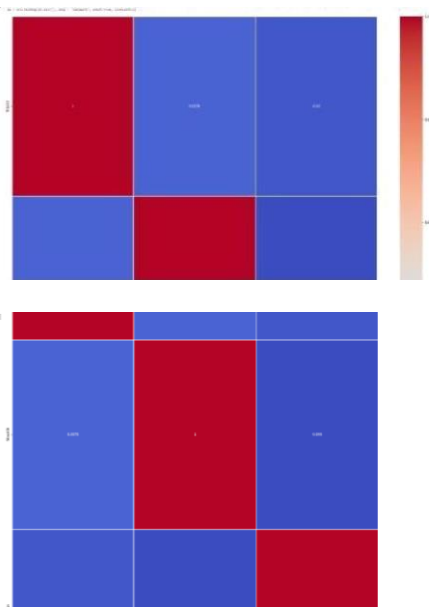
{x}

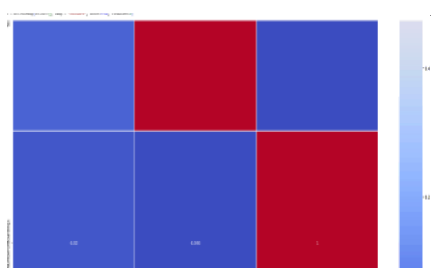
11A Marion Rd	2527
23 Findon Rd	2319
219 Woodville &d	2283
L7 Grange Rd	2093
220 Woodville Rd	2092
15 Portrush Rd	2
L48 Portrush id	1
UI Portrush id	1

Wave: StQpAame, Lesgfn: 92, dbype: i<t64

[28] ## Correlation
plt.figure(figsize=(25,25))
ax = sns.heatmap(df.corr(), cmap = "coolwarm", annot=True, linewidth=2)

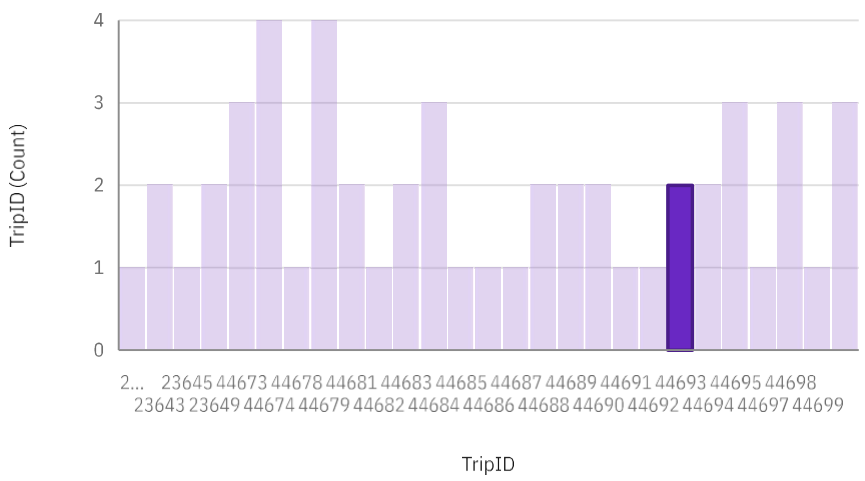
‹ioytbon- input-78-6eb7a2dfbJ3e›:3. FuturWarning: The defaul t value of nuer ic_onLy in Datafrazie.corr is deprecated. In a fatnre '/erszon, zt yi\L defau \t to False. Sel
ax = s n s. heatmap(df. corf'( ) , chap = "coolWa re", annot=Twe, I znevidth -2)



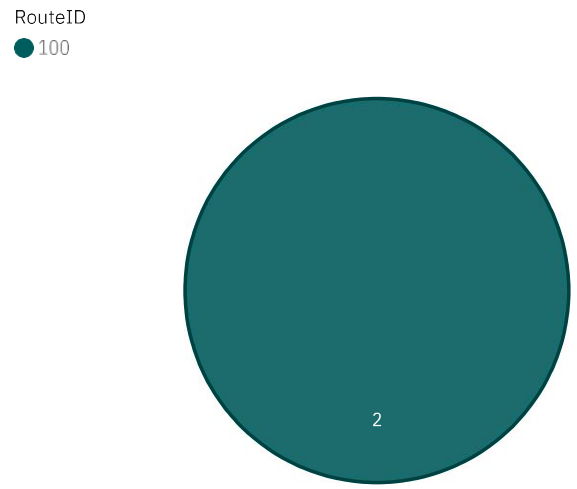


Tab 1

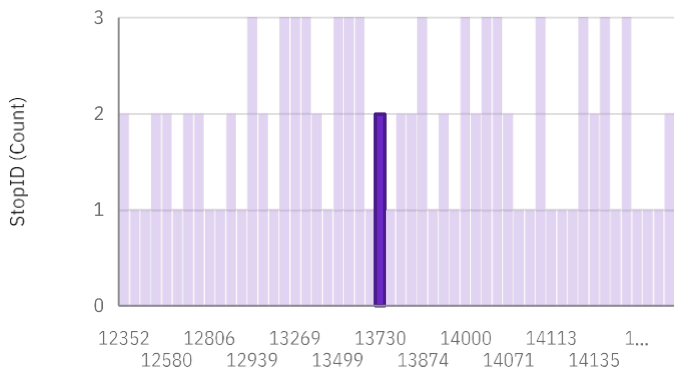
TripID by TripID



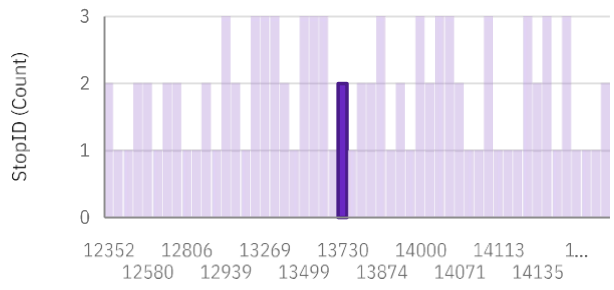
RouteID by RouteID

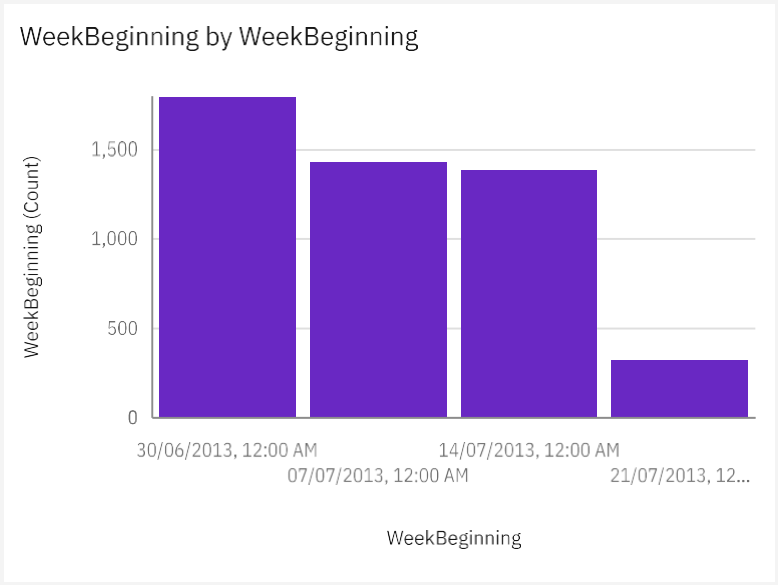


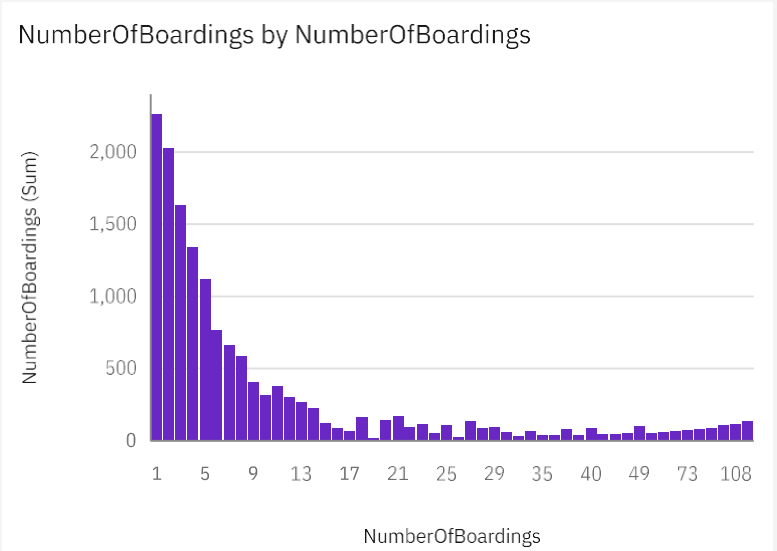
StopID by StopID



StopID by StopID




WeekBeginning by WeekBeginning
WeekBeginning (Count)
1,500
1,000
500
0
30/06/2013, 12:00 AM
07/07/2013, 12:00 AM
14/07/2013, 12:00 AM
21/07/2013, 12...
WeekBeginning


NumberOfBoardings by NumberOfBoardings
NumberOfBoardings (Sum)
2,000
1,500
1,000
500
0
1
5
9
13
17
21
25
29
35
40
49
73
108
NumberOfBoardings

Summary:

wesummarizethekeyfindings,insights,andachievementsofourentireproject. Thissectionservesasaculminatingperspective,bringingtogetherthevariousphasesandeffortsundertakentoaddressthechallengesand opportunities in public transportation.

Inthispartyouwillcontinuebuildingyour project.

Continue building the analysis by creating visualizations using IBM Cognos and integrating code for data analysis.

Design dashboards and reports inIBM Cognos to visualize on-time performance, passenger feedback, and service efficiency metrics. Usecode(e.g., Python)toperformadvanceddataanalysis, suchascalculatingservicepunctualityrates orsentimentanalysison passenger feedback.

Throughout the project, we explored the public bus transport dataset, from its initial design and problem definition to the practical implementationofdatapreprocessing,analysis,andmodeling.Wetackledreal-worldissuesrelatedtopublictransportationefficiencyand strove to find actionable solutions.

Inthis section, we encapsulate the journey by highlighting the project's keyoutcomes, the lessons learned, and any actionable recommendationsderivedfromouranalysis.Thisconclusionsignifiesthesuccessfulcompletionofourpublictransportationefficiency analysis and paves the way for practical applications and informed decisions in the realmofpublic transportation.