Code files are arranged in folders by the section(s) they were used for. Steps refer to the order in which the code files were created and used in the project.

[**1. Introduction**](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.pxix8nub3it)

[**2. Data Cleaning**](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.s7btuwov4yxc)

[2.1 Downloading Data](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.mtnrz4nqgrhv)

*Step 1: 2.1Data Download.sh*

[2.2 Airlines Data](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.5tw09yaqzxuz)

*Step 1: 2.2Create\_Hive\_Table.hive*

*Step 2: 2.2MissingValues.sh*

*Step 3: 2.2HiveAnalysis.hive*

*Step 4: 2.2DivertedMissing.sh*

*Step 5: 2.2DivertedAnalysis.hive*

*Step 6: 2.2Percentile.sh*

*Step 7: FinalData.hive*

(Please also find attached an excel file containing Exploratory Data Analysis Findings named EDA.xlsx)

[2.3 Airports Data](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.2wy6oqiz8ix)

*See code in 4.2.2*

[2.4 Plane Data](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.3j7wyolv6efb)

*See code in 4.2.1*

[**3. Preliminary Analysis**](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.d3khlimp7cm8)

[3.1 Trend in Overall Number of Flights Across Years](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.8gcmkjpkrj3o)

[3.2 Trend in Overall Number of Flights Across Months](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.ogo672kqapib)

*Step 1: 3.1DataExtract.sh*

*Step 2: 3.1Trends.r*

[3.3 Trend in Overall Number of Flights Across Days of the Week](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.fiq6uh5mibhd)

*Step 1:3.1Trends.r*

[3.4 Major Characteristic Comparisons Among States](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.79zmy2t6io1a)

*Step 1: 3.4 heatmap analysis.sh*

*Step 2: 3.4 heatmap analysis.Rmd*

[3.5 Trends in Flight Movement Within and Outside Regions](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.96z3ed161e2t)

*Step 1: 3.5ChordDiagram.sh*

*Step 2: 3.5ChordDiagram.r*

[**4. Analysis**](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.vzsu2wa41zxg)

[4.1 Cancellation Rates](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.40wui6nych65)

[4.1.1 Preliminary Analysis : Comparison of Monthly Cancellation Rates (2000 vs. 2003)](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.btmdtsbvpqak)

*Step 1: 4.1.1RegionalCancellationRate.sh*

*Step 2: 4.1CancellationRate.r*

[4.1.2 Comparison of Cancellation Rates by Carrier (2000 vs. 2003)](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.dma988bxysfw)

*Step 1: 4.1.2, 4.2.3 treemap table creation.sh*

*Step 2: 4.1.2 treemap for cancellation rate.Rmd*

[4.1.3 Comparison of Monthly Cancellation Rates by Carrier (2000 vs. 2003)](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.9siro88rhgma)

*Step 1: 4.1.3 streammap table creation.sh*

*Step 2: 4.1.3 streammap for cancellation by carrier.Rmd*

[4.1.4 Comparison of Monthly Cancellation Rates by Region (2000)](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.4gvl96h94wq0)

*Step 1: 4.1CancellationRate.r*

[4.1.5 Comparison of Monthly Cancellation Rates by Region (2003)](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.xea272tmkmo)

*Step 1: 4.1CancellationRate.r*

[4.1.6 Cancellation Rate by Manufacturer (2000 vs. 2003)](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.rty0je7oft9p)

*Step 1: 4.1.6cancellation\_manufacturer.sh*

Step 2: *4.1.6cancellation\_manufacturer.r*

[4.2 Flight Delays](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.kwiqvbmr6u9)

[4.2.1 Trends of Arrival Delay](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.ttoauscwm7g3)

*Step 1: 4.1.2, 4.2.3 treemap table creation.sh*

*Step 2: 4.2.2 Trends of Arrival Delay.Rmd*

[4.3 Diverted Flights](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.s4dwq3yckl6t)

[4.3.1 Trends in Number of Diverted Flights Across Time](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.ppo4g28woufg)

[4.3.2 Trends in Diverted Flights by Destination](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.svvwwm5ewcx6)

[4.3.3 Top 10 Routes with the Most Number of Diverted Flights](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.mipiwnlr89nv)

*Codes for the entire section:*

*Step 1: 4.3Diverted\_Flight.sh*

*Step 2: 4.3Diverted\_Flight.r*

[4.4 Flight Paths and Number of Flights](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.cpgm9ggfxurq)

[4.4.1 Mapping Flight Paths](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.bv88g3f5ro0c)

*Step 1: 4.4.1MakeMapTable.hive*

*Step 2: 4.4.1MakeMapTable2.hive*

*Step 3: 4.4.1MakeMapTable3.hive*

*Step 4: 4.4.1MapFlights.R*

[4.4.2 Number of Flights By Airport and State](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.69bg3qochkh8)

*Step 1: 4.4.2MakeTreemapTable.hive*

*Step 2: 4.4.2MakeTreemaps.R*

[4.5 Top 10 Airplane Models Used for Long and Short Duration Flights](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.az6f6lhzax33)

*Step 1: 4.5flight\_model\_vs\_elaspe\_time.sh*

*Step 2: 4.5Flight\_duration\_treemaps.Rmd*

[4.5.1 Long and Short Duration Airplane Models Overview](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.t9eo2rkh2wt5)

[4.5.2 Long and Short Duration Airplane Models in 2000 vs. 2003](https://docs.google.com/document/d/1cndbls7883SyVsNDeCVpSfHVAeYvPT050Vfp5KZxGvU/edit#heading=h.t5a8mv5xx6iz)