For this assignment, you’ll need to write a 1000-word report on your exploratory analysis of the following data set. Your report should detail your process and present a summary of your initial findings and insights.

Step 1: Read the business scenario

You work for a company that specialises in last-minute weekend flights in the USA. Your manager has asked you to investigate historical data for domestic flight delays departing from Los Angeles International (LAX) in 2015 to help drive strategic and operational decision-making.

Specifically, the company is looking to use data to help inform decision-making around:

* **Common reasons for delays.**Are there common types of delays for certain airports/areas/states (or by airline)? What about in certain months?
* **Advice to give passengers.**
* **Miscellaneous areas to improve.** Are there ways to improve quality, customer experience, reduce costs, reduce landing/departing bottlenecks?

Your manager suggested that you should check the datasets for errors. He also recommended that you should investigate the flight data using Excel (or similar), before loading it into an SQLite database to run your queries. Any rows with calculation errors involving time points should be deleted, and any calculation errors involving only time durations should be rectified.

Step 2: Access the dataset

Select the following to download the folder which contains flight datasets with errors:

* [Flight datasets [ZIP 195.2MB]](https://swinburneonline.instructure.com/courses/3883/files/2707130?wrap=1)[Download Flight datasets [ZIP 195.2MB]](https://swinburneonline.instructure.com/courses/3883/files/2707130/download?download_frd=1)

**About the data**

You have been provided with three data sets. The first contains details of domestic flights, departing from LAX in 2015. You have also been provided reference data sets for US airports and US airlines.

The flight data set has several columns including the following calculated, or coded columns:

* DAY\_OF\_WEEK - represents scheduled day of the week for the flight, where 1 = Monday, 2 = Tuesday etc..
* Airline, origin and destination airports use the relevant IATA identifiers
* TAXI\_OUT - The time duration elapsed between departure from the origin airport gate and wheels off
* WHEELS\_OFF - The time point that the aircraft's wheels leave the ground
* SCHEDULED\_TIME - Planned time amount needed for the flight trip
* AIR\_TIME - The time duration between wheels\_off and wheels\_on time
* DISTANCE - Distance between two airports
* WHEELS\_ON - The time point that the aircraft's wheels touch on the ground
* TAXI\_IN - The time duration elapsed between wheels-on and gate arrival at the destination airport
* SCHEDULED\_ARRIVAL - Planned arrival time

The following columns were calculated automatically:

* DEPARTURE\_TIME = WHEEL\_OFF - TAXI\_OUT
* ARRIVAL\_TIME = WHEELS\_ON+TAXI\_IN
* ARRIVAL\_DELAY = ARRIVAL\_TIME-SCHEDULED\_ARRIVAL
* ELAPSED\_TIME = AIR\_TIME+TAXI\_IN+TAXI\_OUT

**Note:** all time durations are stored as integers and represent a total number of minutes (i.e. 120 = 2 hours). All time points are stored as integers but represent times in the 24h clock (i.e. 5 = 00:05, 2259 = 22:59, 2260 is invalid).

The airline data set includes two columns:

* IATA\_CODE - International Air Transport Association’s airline identifiers
* AIRLINE - airline’s name

The airport data set has several columns including:

* IATA\_CODE - International Air Transport Association’s location identifier
* AIRPORT - airport name

Step 3: Complete the exploratory analysis

Make sure you use the spreadsheet and database tools in the unit and refer to the examples and supporting material to complete components of your exploratory analysis.

* In [Week 2](https://swinburneonline.instructure.com/courses/3883/pages/2-dot-1-this-weeks-focus?module_item_id=303820) you became familiar with calculating different statistical measures, including measures of central value and measures of spread.
* In [Week 3](https://swinburneonline.instructure.com/courses/3883/pages/3-dot-1-this-weeks-focus?module_item_id=303829) you explored working with multiple datasets.
* In [Week 4](https://swinburneonline.instructure.com/courses/3883/pages/4-dot-1-this-weeks-focus?module_item_id=303837) you applied diagnostic analysis techniques to identify potential causes of irregularities.

To complete the exploratory analysis, you should:

* identify and extract relevant fields from the data set
* perform any necessary cleaning or transformation of the data
* create exploratory visualisations or calculate relevant summary statistics to identify patterns, trends, or insights.

As you gain insight, you might want to repeat the exploratory process, including more data or using different descriptive or diagnostic techniques as required.

 Step 4: Complete your report

In your report, you should:

* summarise the business problem and identify data to be used in your analysis
* summarise the steps taken to prepare data for analysis, including your approach to identify and manage outliers
* describe any transformation of the original data, combination of data sources, or the inclusion of external data sources as required
* justify the choice of exploratory visualisations, including examples where appropriate
* justify your choice of descriptive measures and/or diagnostic techniques
* present your initial findings and insights from the given data set.

**Tips for writing the report:**

Your report should meet the following standard:

* Interpreting the business problem
	+ Extensive details of the business problem and selected data have been provided.
	+ Some discussion of the strengths and limitations of the current data, or areas for further investigation provided.
* Describing data wrangling steps
	+ Summary of the steps taken to prepare data for analysis are clear, well organised, appropriate, and include an approach to identify and manage outliers.
	+ Detailed and insightful descriptions of processes and decisions are evident at each stage.
* Identifying patterns and trends
	+ Detailed evidence of exploratory visualisations, descriptive measures or diagnostic techniques has been provided. Extensive examples and detailed written justifications are provided.
* Describing analytic approaches
	+ Detailed and insightful descriptions of processes and decisions made during the analysis are evident at each stage.
	+ Descriptions of any failed experiments, or multiple rounds of analysis are included.
* Communicating findings and insights
	+ Insights and findings are clearly articulated in relation to the original business scenario.

Supporting resources

The following resources will assist you with completing this assignment:

* [Supporting resources](https://swinburneonline.instructure.com/courses/3883/pages/supporting-resources) (Excel, SQLite 3, Tableau, web-based and multimedia resources).
* [Report writing (Links to an external site.)](https://www.swinburne.edu.au/media/swinburneeduau/current-students/docs/pdf/Writing_at_University_2_the_Business_Report.pdf)
* [Sample business report](https://www.swinburne.edu.au/media/swinburneeduau/current-students/docs/pdf/Sample_Text_2_Business_Report_the_Stylish_Report.pdf)

For this assignment, you’ll need to write a 1000-word report on your exploratory analysis of the following data set. Your report should detail your process and present a summary of your initial findings and insights.

Step 1: Read the business scenario

You work for a company that specialises in last-minute weekend flights in the USA. Your manager has asked you to investigate historical data for domestic flight delays departing from Los Angeles International (LAX) in 2015 to help drive strategic and operational decision-making.

Specifically, the company is looking to use data to help inform decision-making around:

* **Common reasons for delays.**Are there common types of delays for certain airports/areas/states (or by airline)? What about in certain months?
* **Advice to give passengers.**
* **Miscellaneous areas to improve.** Are there ways to improve quality, customer experience, reduce costs, reduce landing/departing bottlenecks?

Your manager suggested that you should check the datasets for errors. He also recommended that you should investigate the flight data using Excel (or similar), before loading it into an SQLite database to run your queries. Any rows with calculation errors involving time points should be deleted, and any calculation errors involving only time durations should be rectified.

Step 2: Access the dataset

Select the following to download the folder which contains flight datasets with errors:

* [Flight datasets [ZIP 195.2MB]](https://swinburneonline.instructure.com/courses/3883/files/2707130?wrap=1)[Download Flight datasets [ZIP 195.2MB]](https://swinburneonline.instructure.com/courses/3883/files/2707130/download?download_frd=1)

**About the data**

You have been provided with three data sets. The first contains details of domestic flights, departing from LAX in 2015. You have also been provided reference data sets for US airports and US airlines.

The flight data set has several columns including the following calculated, or coded columns:

* DAY\_OF\_WEEK - represents scheduled day of the week for the flight, where 1 = Monday, 2 = Tuesday etc..
* Airline, origin and destination airports use the relevant IATA identifiers
* TAXI\_OUT - The time duration elapsed between departure from the origin airport gate and wheels off
* WHEELS\_OFF - The time point that the aircraft's wheels leave the ground
* SCHEDULED\_TIME - Planned time amount needed for the flight trip
* AIR\_TIME - The time duration between wheels\_off and wheels\_on time
* DISTANCE - Distance between two airports
* WHEELS\_ON - The time point that the aircraft's wheels touch on the ground
* TAXI\_IN - The time duration elapsed between wheels-on and gate arrival at the destination airport
* SCHEDULED\_ARRIVAL - Planned arrival time

The following columns were calculated automatically:

* DEPARTURE\_TIME = WHEEL\_OFF - TAXI\_OUT
* ARRIVAL\_TIME = WHEELS\_ON+TAXI\_IN
* ARRIVAL\_DELAY = ARRIVAL\_TIME-SCHEDULED\_ARRIVAL
* ELAPSED\_TIME = AIR\_TIME+TAXI\_IN+TAXI\_OUT

**Note:** all time durations are stored as integers and represent a total number of minutes (i.e. 120 = 2 hours). All time points are stored as integers but represent times in the 24h clock (i.e. 5 = 00:05, 2259 = 22:59, 2260 is invalid).

The airline data set includes two columns:

* IATA\_CODE - International Air Transport Association’s airline identifiers
* AIRLINE - airline’s name

The airport data set has several columns including:

* IATA\_CODE - International Air Transport Association’s location identifier
* AIRPORT - airport name

Step 3: Complete the exploratory analysis

Make sure you use the spreadsheet and database tools in the unit and refer to the examples and supporting material to complete components of your exploratory analysis.

* In [Week 2](https://swinburneonline.instructure.com/courses/3883/pages/2-dot-1-this-weeks-focus?module_item_id=303820) you became familiar with calculating different statistical measures, including measures of central value and measures of spread.
* In [Week 3](https://swinburneonline.instructure.com/courses/3883/pages/3-dot-1-this-weeks-focus?module_item_id=303829) you explored working with multiple datasets.
* In [Week 4](https://swinburneonline.instructure.com/courses/3883/pages/4-dot-1-this-weeks-focus?module_item_id=303837) you applied diagnostic analysis techniques to identify potential causes of irregularities.

To complete the exploratory analysis, you should:

* identify and extract relevant fields from the data set
* perform any necessary cleaning or transformation of the data
* create exploratory visualisations or calculate relevant summary statistics to identify patterns, trends, or insights.

As you gain insight, you might want to repeat the exploratory process, including more data or using different descriptive or diagnostic techniques as required.

 Step 4: Complete your report

In your report, you should:

* summarise the business problem and identify data to be used in your analysis
* summarise the steps taken to prepare data for analysis, including your approach to identify and manage outliers
* describe any transformation of the original data, combination of data sources, or the inclusion of external data sources as required
* justify the choice of exploratory visualisations, including examples where appropriate
* justify your choice of descriptive measures and/or diagnostic techniques
* present your initial findings and insights from the given data set.

**Tips for writing the report:**

Your report should meet the following standard:

* Interpreting the business problem
	+ Extensive details of the business problem and selected data have been provided.
	+ Some discussion of the strengths and limitations of the current data, or areas for further investigation provided.
* Describing data wrangling steps
	+ Summary of the steps taken to prepare data for analysis are clear, well organised, appropriate, and include an approach to identify and manage outliers.
	+ Detailed and insightful descriptions of processes and decisions are evident at each stage.
* Identifying patterns and trends
	+ Detailed evidence of exploratory visualisations, descriptive measures or diagnostic techniques has been provided. Extensive examples and detailed written justifications are provided.
* Describing analytic approaches
	+ Detailed and insightful descriptions of processes and decisions made during the analysis are evident at each stage.
	+ Descriptions of any failed experiments, or multiple rounds of analysis are included.
* Communicating findings and insights
	+ Insights and findings are clearly articulated in relation to the original business scenario.

Supporting resources

The following resources will assist you with completing this assignment:

* [Supporting resources](https://swinburneonline.instructure.com/courses/3883/pages/supporting-resources) (Excel, SQLite 3, Tableau, web-based and multimedia resources).
* [Report writing (Links to an external site.)](https://www.swinburne.edu.au/media/swinburneeduau/current-students/docs/pdf/Writing_at_University_2_the_Business_Report.pdf)
* [Sample business report](https://www.swinburne.edu.au/media/swinburneeduau/current-students/docs/pdf/Sample_Text_2_Business_Report_the_Stylish_Report.pdf)
* Interpret the business problem.
* Describe data wrangling steps.
* Identify patterns and trends.
* Describe analytical approach.
* Communicate findings and insights.

Your work will be assessed using the following marking guide:

|  |
| --- |
|  |

| *Assignment 1 marking guide* |
| --- |
| **Criteria** | **No Pass** | **Pass****50–59%** | **Credit****60–69%** | **Distinction****70–79%** | **High Distinction****80–100%** |
| Interpret the business problem.(20%) | Did not meet criterion. | Brief or partial details of the business problem and selected data have been provided.  | Details of the business problem and selected data have been provided.  | Extensive details of the business problem and selected data have been provided.   | Extensive details of the business problem and selected data have been provided. Some discussion of the strengths and limitations of the current data, or areas for further investigation provided. |
| Describe data wrangling steps.(20%) | Did not meet criterion. | Partial summary of the steps taken to prepare data for analysis are evident. There may not be a clear approach to identify and manage outliers. Descriptions of process and decisions are evident at some stages but not all. | Summary of the steps taken to prepare data for analysis are clear, appropriate, and include an approach to identify and manage outliers. Clear descriptions of process and decisions are evident at each stage. | Summary of the steps taken to prepare data for analysis are clear, well organised, appropriate, and include an approach to identify and manage outliers. Strong descriptions of process and decisions are evident at each stage. | Summary of the steps taken to prepare data for analysis are clear, well organised, appropriate, and include multiple approaches to identify and manage outliers. Detailed and insightful descriptions of process and decisions are evident at each stage. |
| Identify patterns and trends.(20%) | Did not meet criterion. | Partial evidence of exploratory visualisations, descriptive measures or diagnostic techniques have been provided. Examples and/or brief justifications are included. | Clear evidence of exploratory visualisations, descriptive measures or diagnostic techniques have been provided. Examples and written justifications are appropriate and documented. | Clear evidence of exploratory visualisations, descriptive measures or diagnostic techniques have been provided. Strong examples and written justifications are provided. | Detailed evidence of exploratory visualisations, descriptive measures or diagnostic techniques have been provided. Extensive examples and detailed written justifications are provided. |
| Describe analytical approach.(20%) | Did not meet criterion. | Partial descriptions of process and decisions made during the exploratory analysis are evident at some stages but not all. | Clear descriptions of process and decisions made during the exploratory analysis are evident at each stage. | Detailed descriptions of process and decisions made during the exploratory analysis are evident at each stage. | Detailed and insightful descriptions of process and decisions made during the analysis are evident at each stage.Descriptions of any failed experiments, or multiple rounds of analysis are included. |
| Communicate findings and insights.(20%) | Did not meet criterion. | Partial summary of the findings has been provided.  | Clearly articulated findings in relation to the original business scenario. | Clearly articulated findings and has provided some good insights in relation to the original business scenario. |  Detailed and extensive insights and findings in relation to the original business scenario. |