

FIT3003 Assignment 2 - S2 2024 (Weight = 40%)
Due - Friday, 11 October 2024, 4:30 PM

Version: 3.0 – 17/09/2024

General Information and Submission

- This is an individual assignment.
- *Submission method:* Submission is online through Moodle.
- *Penalty for late submission:* 5% deduction for each day.
- *Assignment FAQ:* There is an Assignment Frequently Asked Questions page set up for the Assignment 2 on EdStem Forum.

Problem Description

M-Stay is a residential service that offers homestay and rental services to Monash students and staff around Melbourne. The company has an existing operational database that maintains and stores all of the business transactions information (e.g., properties, hosts, listings, booking, etc.) required for the management's daily operation. As the business grows, M-Stay has decided to build a Data Warehouse to improve their analysis and work efficiency. However, since the staff at M-Stay have limited Business Intelligence and Data Warehouse knowledge, they have decided to hire you to design, develop and quickly generate BI reports from a Data Warehouse.

The operational database tables can be found at the [MStay](#) account. You can, for example, execute the following query:

select * from MStay.<table_name>;

The data definition of each table in MStay is as follows:

Table Name	Attributes,Data Types and Key Constraints		Notes
REVIEW	Review_ID	Number (PK)	The table stores review information of the related booking order.
	Review_Date	Date	
	Review_Comment	Varchar2	
	Booking_ID	Number (FK)	

BOOKING	Booking_ID	Number (PK)	The table stores booking information.
	Booking_Date	Date	
	Booking_Stay_Start_Date	Date	
	Booking_Duration	Number	
	Booking_Cost	Number	
	Booking_Num_Guests	Number	
	Listing_ID	Number (FK)	
	Guest_ID	Number (FK)	
GUEST	Guest_ID	Number (PK)	The table stores all guest information.
	Guest_Name	Varchar2	
LISTING	Listing_ID	Number (PK)	The table stores all listing information. Each listing has one property and one host information.
	Listing_Date	Date	
	Listing_Title	Varchar2	
	Listing_Price	Number	
	Listing_Min_Nights	Number	
	Listing_Max_Nights	Number	
	Prop_ID	Number (FK)	
	Type_ID	Number (FK)	
	Host_ID	Number	

		(FK)	
HOST	Host_ID	Number (PK)	The table stores all host information.
	Host_Name	Varchar2	
	Host_Since	Date	
	Host_Location	Varchar2	
	Host_About	Varchar2	
	Host_Listing_Count	Number	
HOST_VERIFICATION	Host_ID	Number (PF)	The table stores the verification information between host and channel.
	Channel_ID	Number (PF)	
CHANNEL	Channel_ID	Number (PK)	The table stores the channel of verification for the hosts.
	Channel_Name	Varchar2	
LISTING_TYPE	Type_ID	Number (PK)	The table stores all listing types.
	Type_Description	Varchar2	
PROPERTY	Prop_ID	Number (PK)	The table stores all property information.
	Prop_Description	Varchar2	
	Prop_Neighbourhood_Overview	Varchar2	
	Prop_Num_Beds	Number	
	Prop_Num_Bedrooms	Number	
	Prop_Num_Bathrooms	Number	
	Prop_Num_Reviews	Number	

	Prop_Rating_Location	Number	
	Prop_Rating_Cleanliness	Number	
	Prop_Rating_Value	Number	
	Prop_Average_Rating	Number	
PROPERTY_AMENITY	Prop_ID	Number (PF)	The table links property and amenity tables
	Amm_ID	Number (PF)	
AMENITY	Amm_ID	Number (PK)	The table stores all amenities information
	Amm_Description	Varchar2	

Assignment Project Exam Help

A. Transformation Stage

The first stage of this assignment is divided into **TWO** main tasks:

1. Design a data warehouse for the above M-Stay database.

You are required to create a data warehouse for the **M-Stay** database.

The management is especially interested in the following indicators:

- Number of reviews
- Number of listings
- Average booking cost (find appropriate fact measures that can calculate the average booking cost)

The following is a list of dimension attributes that you should include in your data warehouse:

- Listing type
- Listing time [Month, Year]
- Listing season
 - (Spring: 9 to 11, Summer: 12 to 2, Autumn: 3 to 5 and Winter: 6 to 8)
- Listing maximum stay duration [*short-term*: less than 14 nights, *medium-term*: 14 to 30 nights, *long-term*: more than 30 nights]

- Listing price range [*low*: less than \$100, *medium*: \$100 to \$200, *high*: more than \$200]
- Channels
- Booking duration [*short-term*: less than 30 nights, *medium-term*: 30 to 90 nights, *long-term*: more than 90 nights]
- Review time [Month, Year]
- Booking cost range [*low*: less than \$5000, *medium*: \$5000 to \$10000, *high*: more than \$10000]

For the attribute, ensure that it meets the requirements of the range or group specified in your submission, if required in the specification.

- **Preparation stage.**

Before you start designing the data warehouse, you have to ensure that you have explored the operational database and have done sufficient data cleaning. Once you have done the data cleaning process, you are required to explain what strategies you have taken to explore and clean the data.

The outputs of this task for **Report** are:

- a) If you have done the data cleaning process, explain the strategies you used in this process (you need to show the SQL to explore the operational database and SQL of the data cleaning, as well as the screenshot of data *before* and *after* data cleaning).

Add WeChat powcoder

- **Designing the data warehouse by drawing star/snowflake schema.**

Design task A:

The star schema for this data warehouse may contains multi-facts. You need to identify the fact measures, dimensions, and attributes of the star/snowflake schema. The following queries might help you to determine the fact measures and dimensions:

- How many long-term stay duration listings are listed on Facebook?
- How many listings are listed in June 2015?
- How many listings are there in summer for an “Entire home/apt” in a medium price range?
- How much is the average booking cost in March 2013?
- How many bookings were there for “Private rooms” with a short-term stay duration in 2015?
- How many high-cost bookings were made in April 2014?
- How many reviews were given in February 2016?

Note: the star schema you created in Design Task A as the highest level of aggregation

Design task B:

In this assignment, consider the star schema you created in Design Task A as the highest level of aggregation. The M-Stay company manager wants to implement a drill-down function to explore more detailed information. Your task is to **suggest several ways** to increase the granularity of your fact tables from Design Task A. In other words, the manager wants to decrease the aggregation level of the fact tables you created in Design Task A.

The outputs of task A & B for **Report** are:

- b) A star/snowflake schema diagrams for design task A (You can use [Lucidchart](#) to draw the star schema).
- c) List suggestion of increase the granularity of your fact tables for design task B.

2. Implement **design task A** star/snowflake schema using SQL.

You are required to implement the star/snowflake schema that you have drawn in **design task A**. This implies that you need to create the fact and dimension tables in SQL. The output is a series of SQL statements to perform this task. You will also need to show that this task has been carried out successfully.

Note:

- If your account is full, you will need to drop all of the tables that you have previously created during the tutorials.
- If you have dropped all tables in your account and you still encounter the `ORA-01536: space quota exceeded for tablespace 'TABLE_NAME'`, please check your SQL code whether you have properly joined all tables. This issue was mainly caused when you did not do the table join properly as the number of records multiplied during the process.

The outputs of this task for **Report** are:

- a) Screenshots of the table structure you created for Design Task A, including the dimension table and fact tables.

A sample of screenshots of the table structure

REVIEW					
Columns Data Model Constraints Grants Statistics Triggers Flashback Dependencies Details Partitions Indexes SQL					
Actions...					
	COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID COMMENTS
1	REVIEW_ID	NUMBER(10,0)	Yes	(null)	1 (null)
2	REVIEW_DATE	DATE	Yes	(null)	2 (null)
3	REVIEW_COMMENT	VARCHAR2(4000 BYTE)	Yes	(null)	3 (null)
4	BOOKING_ID	NUMBER(15,0)	Yes	(null)	4 (null)

B. Data Analytic Stage

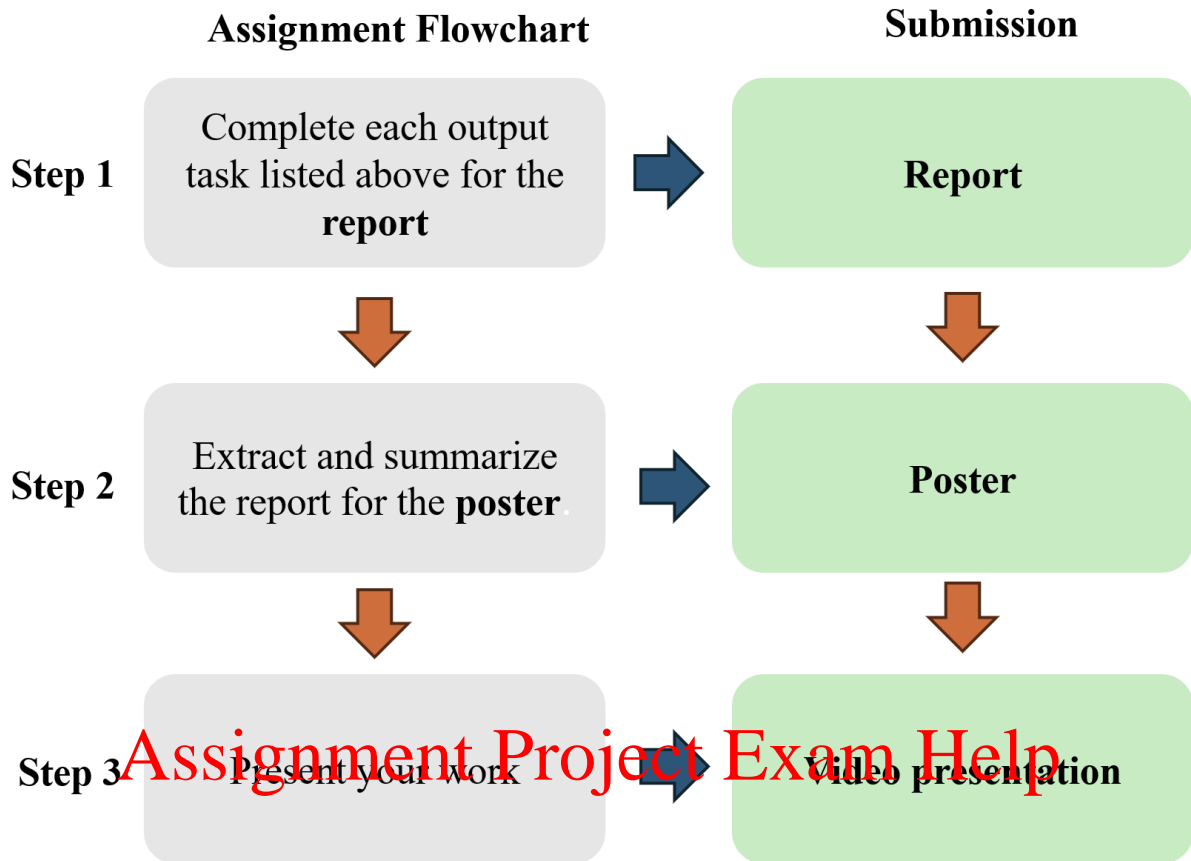
Conduct a data analysis using the star schema you created in Design Task A by writing SQL queries to explore the data further. Present your findings in a clear and concise manner, demonstrating your understanding of the dataset and highlighting any noteworthy observations or patterns.

The outputs of this task for **Report** are:

Findings report: A detailed explanation of your findings, including any significant observations or patterns identified during the analysis.

Submission Checklist

<https://powcoder.com>
Add WeChat powcoder



Assignment Project Exam Help
<https://powcoder.com>

Add WeChat powcoder

Step 1: Report (25% of the total score)

A combined pdf file save as: YourstudentID_A2_report.pdf, containing all of the above tasks:

- A. Cover page
- B. If you have done the data cleaning process, explain the strategies you used in this process (you need to show the SQL to explore the operational database and SQL of the data cleaning, as well as the screenshot of data *before* and *after* data cleaning). Note that you are only required to find around 5 (five) data errors for this stage.
- C. A star/snowflake schema diagrams for design task A
- D. List suggestion of increase the granularity of your fact tables for design task B
- E. Screenshots of the table structure you created for Design Task A only, including the dimension table and fact tables.
 - a. SQL file for creating the star schema is **NOT** required in submission
- F. Findings report: A detailed explanation of your findings, including any significant observations or patterns identified during the analysis.

Step 2: Poster (35% of the total score)

One page standard A4 poster in PDF format to save as: YourstudentID_A2_poster.pdf

Extract key information from the report you created and present it in a **one-page poster**. The poster must be in **standard A4 size and in PDF format**, which can be either landscape or portrait. The content should be clear and easy to understand. Avoid using technical jargon or complex language. Review the poster before submission to ensure it effectively communicates the key messages of your report.

Note:

Ensure the poster content is consistent with the key structure and findings of your report, and choose an appropriate layout that effectively organizes the information in a clear and logical manner. Maintain a good balance of text and visuals to enhance readability, and ensure all visuals are relevant and support the content of the poster. Label all visuals clearly and provide captions where necessary. Avoid overcrowding the poster with too much text or too many visuals, and ensure the poster is free of any grammatical or typographical errors.

Key guidance of design a poster:

- What is the main theme/objective of the poster that you want to express?
- Who is your target audience for this poster?
- Do you really need all the details from your report on this poster?

Step 3: Video presentation (40% of the total score)

A five minute video presentation in mp4 format save as: YourstudentID_A2_video.mp4

Based on the report and poster you have created, present your design and findings in a **five-minute video presentation**. Ensure you thoroughly understand both the report and the poster to effectively extract and communicate the key points.

Assignment Submission

The assignment must be submitted electronically through Moodle. Please ensure the following:

1. Step 1 output: A combined pdf file save as: YourstudentID_A2_report.pdf
2. Step 2 output: One page standard A4 poster in PDF format to save as: YourstudentID_A2_poster.pdf
3. Step 3 output: A five minute video presentation in mp4 format save as: YourstudentID_A2_video.mp4

Zip all above files from step 1 to 3, and name the ZIP folder as A2_YourstudentID.zip.

The due date: **Due - Friday, 11 October 2024, 4:30 PM**

- **Assignment Project Exam Help**
The Submission of this assignment must be in the form of a single ZIP file. Only PDF and .mp4 files will be accepted within the zip file. No other formats will be accepted.
- **https://powcoder.com**
You must ensure that you have all the files listed in this checklist before submitting your assignment to Moodle. Failure to submit a complete list of files will lead to mark penalties.
- **Add WeChat powcoder**
It is important to note that our support hours are limited and we don't have the capacity to deal with submission issues outside of working hours. You must ensure that you have all the files listed in this checklist before submitting your assignment to Moodle. Failure to submit a complete list of files will result in a mark penalty.
- Penalty for late submission: 5 % deduction for each day, including weekends.
- Submission Cut-off time: **18 October 2024, 4:30 PM** (Submission link will be unavailable after the cut-off date).

Authorship

This assignment is an **individual assignment** and the final submission must be identifiably your own work. Breaches of this requirement will result in an assignment not being accepted for assessment and may result in disciplinary action.

As per the University's policy on the guidelines and practice pertaining to the usage of Generative AI, this assignment restricts all use of generative AI. In this assessment, you must not use generative artificial intelligence (AI) to generate any materials or content in relation to the assessment task.

Late Penalty:

Late assignments submitted without an approved extension may be accepted up to a maximum of **seven days** with the approval of the Chief Examiner and/or Lecturer but will be **penalised at the rate of 5% per day (including weekends and public holidays)**. Assignments submitted more than seven days after the due date will receive a zero mark for that assignment and may **not receive any feedback**.

Please note:

- An inability to manage your time or computing resources will not be accepted as a valid excuse. (Several assignments being due at the same time are a fact of university life.)
- Hardware failures, whether of personal or university equipment, are not normally recognised as valid excuses. Failure to back up assignment files is also not recognised as a valid excuse.

Special Consideration:

All extensions / special considerations will now be handled by the central Spec Con team.
Please do not email teaching staff to request an extension or special consideration.

Extensions and other individual alterations to the assessment regime will only be considered using the University Special Consideration Policy. Students should carefully read the [Special Consideration website](#), especially the details about what formal documentation is required.

All special consideration requests should be made using the [Special Consideration Application](#).

Please do not assume that submission of a Special Consideration application guarantees that it will be granted – you must receive an official confirmation that it has been granted.

Getting help and support:

What can you get help for?

- **Consultations with the Teaching Team**
Talk to the Teaching Team: <https://lms.monash.edu/course/view.php?id=162086§ion=2>
- **English language skills**
Talk to English Connect: <https://www.monash.edu/english-connect>
- **Study skills**
Talk to a learning skills advisor: <https://www.monash.edu/library/skills/contacts>
- **Counselling**

Talk to a counsellor: <https://www.monash.edu/health/counselling/appointments>

Plagiarism and Collusion:

Monash University is committed to upholding standards and academic integrity and honesty. Please take the time to view these links.

[Academic Integrity Module](#)

[Student Academic Integrity Policy](#)

[Test your knowledge, collusion \(FIT No Collusion Module\)](#)

Assignment Project Exam Help
All the best for your Assignment!

<https://powcoder.com>

Add WeChat powcoder