DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

EXAM ASSIGNMENT

Study Programme and level	**MSc Business Intelligence**						
Term	Winter 2024/25 – ordinary exam						
Course name and exam code(s)	Database Management and Data Visualization				460232E004		
Exam form and duration	Written onsite exam, internet allowed				4 hours		
Date and time	17 December 2024				9:00-13:00		
Supplementary material/aids	All	X	Specified		No		
Anonymous exam	Yes	X	No		Please do **not** write your name or student ID number anywhere.		
Use of generative AI (GAI) allowed	Yes	X	No				
Hand-in of handwritten material allowed	Yes		No	X			
Hand-in of extra material (appendix) in WISEflow allowed	Yes	X	No				
Other relevant information	**Avoid being suspected of exam cheating** Remember to state references and use quotation marks, if you copy text from other sources or re-use parts of a previously submitted exam paper (plagiarism and self-plagiarism). Students must answer the exam assignment **individually**. All submitted exam papers are checked for plagiarism, so cheating and collaboration between students will be detected.						
Number of pages (incl. front page)	6 pages						

Other instructions:
It is important that you start uploading your exam paper well in advance – at least 10 min. before end of exam.

Question 1: Docker (25%)

1.1 Create a new shared volume called “volume_ordinary”. Provide the command for creating the volume as the answer to this subassignment.

In the next couple of subassignments, you are to build an image from the Dockerfile called “docker_file_O_exam” and run a container from this image. The container will run RStudio, and from this RStudio instance, you need to open the file ”run_O_exam.R” and run the lines in this file. Running the “run_O_exam.R” file requires that you also have the script “fun_O_exam.R” in the Docker container.

1.2 Correct the Dockerfile so you can run line 1 in the file “run_O_exam.R”. Copy-paste the corrected Dockerfile in as the answer to this subassignment.

1.3 On your droplet, build an image from “docker_file_O_exam”, and run a container from this image where you attach the shared volume from subassignment 1.1. Provide the command for running the container as the answer to this subassignment.

1.4 Log in to RStudio in your new container, and run the lines in the file “run_O_exam.R”. Provide the output from the R console as the answer to this subassignment.

Question 2: SQL (20%)

2.1 In this subassignment, you should use one or several of the following R functions from class: “psql_append_df”, “psql_manipulate”, or “psql_select”. From RStudio, create an SQL schema called “marketing”, and within this schema, create a table called “marketing_campaign” such that the final table corresponds to the specifications given in Figure 2.1 below. Set a primary key in the table, and ensure it will have autoincrementing integers. Create a data frame in R where the column names are “campaign_name”, “budget”, “start_date”, and “is_active”, respectively. The data frame should have a single row containing the values “Holiday Sale Campaign”, 5000, “2024-12-01 08:00:00.00”, and 1.
Hint: In R, a text string can be transformed into a timestamp format by using for example: *as.POSIXct("2024-12-01 08:00:00.00", format = "%Y-%m-%d %H:%M:%S")*.

Insert this data frame into the SQL table. The answer to this assignment is a screenshot from DBeaver with the final table (holding the data you just inserted) and your R code.

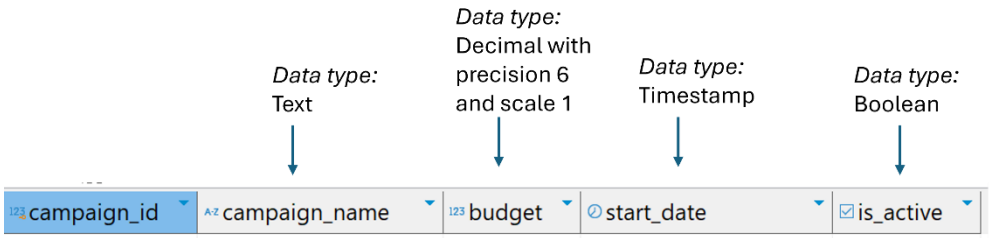

Figure 2.1

2.2 With DBeaver, connect with host: “159.223.30.236”, port: “6432”, database: “exam”, schema: “public”, and your credentials given on Brightspace. For each actor, find the languages of the films they have appeared in. Your query should include all actors and ensure that each actor is listed only once for each language they are associated with. Display the actor’s first name, last name, ID, and the language of the film. Sort the results by actor ID (in ascending order). Provide your query as the answer to this subassignment.

2.3 Create a schema called “bi_trigger”. In this schema, create tables, keys, and insert values following the specifications given in Figure 2.3 below.

When inserting values for *product_sk* in the *product_history* table, use the *default* value. Create a trigger such that updates to *product_name* in the *product_current* table are reflected in the *product_history* table using an appropriate SCD type. Finally, update the product name “Laptop X1 Pro” to “Laptop X1 Pro Business” in the *product_current* table, and verify that the update is reflected in the *product_history* table. As the answers to this subassignment, provide: 1) an explanation of the SCD type you have implemented, and 2) your SQL code.

Schema: *bi_trigger*

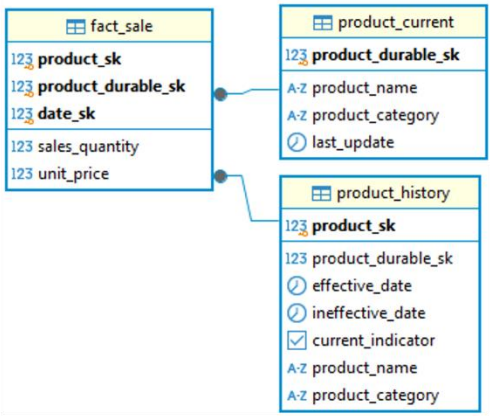

Table: *fact_sale*

product_sk	product_durable_sk	date_sk	sales_quantity	unit_price
1	1	20,230,202	10	349
2	2	20,231,018	5	20
3	1	20,230,915	15	503

Table: *product_history*

product_sk	product_durable_sk	effective_date	ineffective_date	current_indicator	product_name	product_category
1	1	2023-02-01	2023-10-09	[ ]	Laptop X1	Electronics
2	2	2023-10-10	9999-10-10	[v]	Smartphone Z2	Mobile Devices
3	1	2023-11-09	9999-11-09	[v]	Laptop X1 Pro	Electronics

Table: *product_current*

product_durable_sk	product_name	product_category	last_update
1	Laptop X1 Pro	Electronics	2023-11-09
2	Smartphone Z2	Mobile Devices	2023-10-10

Figure 2.3

Question 3: API and Git (25%)

3.1 On rapidapi.com, find and use the API called “MDBList” (no payments necessary). By consulting the API documentation for “MDBList”, construct an API call from R using the httr2 package where the parameter “s” is set to “Christmas”. To answer this subassignment, provide: 1) the R code, and 2) the movie title and ID of the first movie in the response.

3.2 From RStudio in a Docker container, you should first clone the repository *https://github.com/dmdvben/git_2024_1.git*. Next, modify and run the script “run_O_exam2.R”, and provide the console output as the first answer to this subassignment. Lastly, explain what you see in the file “example_conflict.R”, and provide an appropriately modified version of this text file as the second answer to this subassignment.

Question 4: Dashboard (30%)

You are tasked with creating an interactive dashboard in RShiny to explore and visualize weather data.

The dataset contains information about various weather metrics such as wind speeds, pressures, and rain level.

Based on the provided data, you must recreate the shiny app shown in Figures 4.1 and 4.2 below.

The app is based on the dataset provided as a CSV file (exam.csv).

The app consists of two pages:

4.1 A page where the user can filter on Year and Storm Category and select the measures they want to plot. *Hint*: You will have to create a date-time variable for the x axis. (Figure 4.1)

4.2 A page showing summary statistics where the user can select the measures they want a summary on. (Figure 4.2)

To answer this question, you must provide: 1) relevant screenshot(-s) of your dashboard, and 2) the R code in a separate R file.

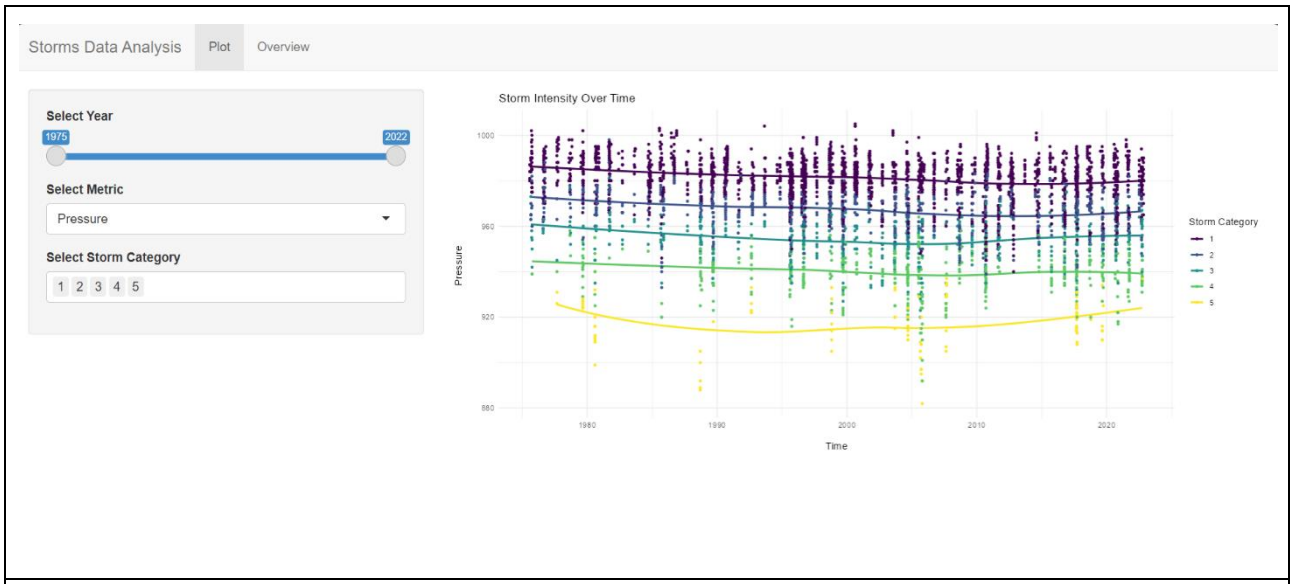

Figure 4.1

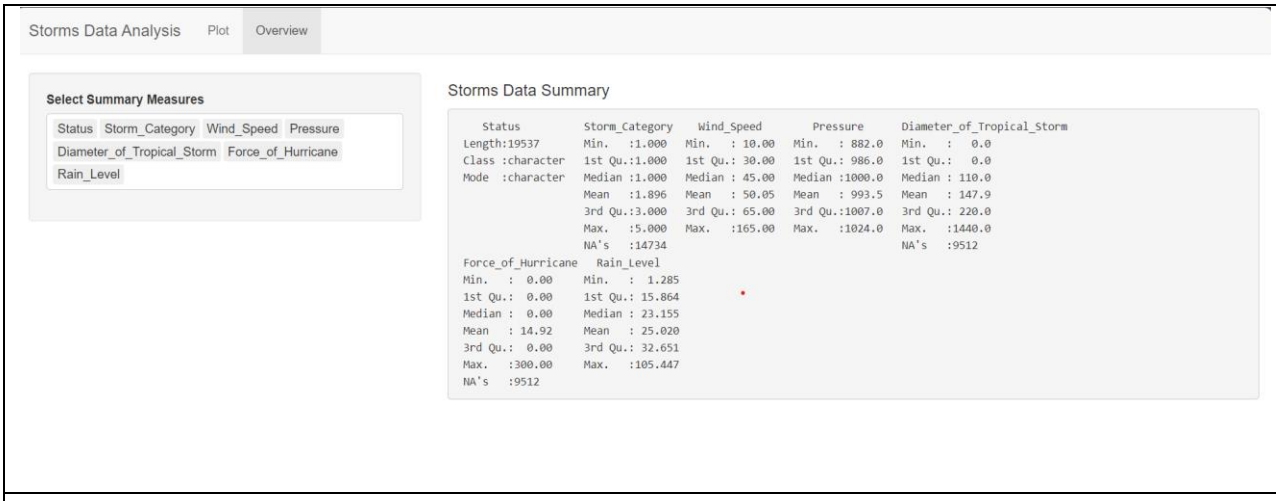
Storms Data Analysis
Plot
Overview
Select Summary Measures
Status
Storm_Category
Wind_Speed
Pressure
Diameter_of_Tropical_Storm
Force_of_Hurricane
Rain_Level
Storms Data Summary
Status Storm_Category Wind_Speed Pressure Diameter_of_Tropical_Storm
Length:19537 Min. :1.000 Min. : 10.00 Min. : 882.0 Min. : 0.0
Class :character 1st Qu.:1.000 1st Qu.: 30.00 1st Qu.: 986.0 1st Qu.: 0.0
Mode :character Median :1.000 Median : 45.00 Median :1000.0 Median : 110.0
Mean :1.896 Mean : 50.05 Mean : 993.5 Mean : 147.9
3rd Qu.:3.000 3rd Qu.: 65.00 3rd Qu.:1007.0 3rd Qu.: 220.0
Max. :5.000 Max. :165.00 Max. :1024.0 Max. :1440.0
NA's :14734 NA's :9512
Force_of_Hurricane Rain_Level
Min. : 0.00 Min. : 1.285
1st Qu.: 0.00 1st Qu.: 15.864
Median : 0.00 Median : 23.155
Mean : 14.92 Mean : 25.020
3rd Qu.: 0.00 3rd Qu.: 32.651
Max. :300.00 Max. :105.447
NA's :9512

Figure 4.2