**School of Computing**

|  |  |
| --- | --- |
| Module Title and Code | BUSINESS INTELLIGENCE - M33150 - FHEQ 7 |
| Module Coordinator  Other lecturers | Dr Elisavet Andrikopoulou (elly.an@port.ac.uk) |
| Assessment Item number | Item 1 |
| Assessment Title | Individual coursework |
| Date Issued | 2024-01-25 |

**Schedule and Deliverables**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Deliverable | Value | Format | Deadline / Date | Late deadline  ECF deadline |
| Individual coursework - presentation | 60% | 15-minute individual presentation + slides | 2024-03-20 16:00 | 2024-04-05 16:00 |

**Notes and Advice**

* The [Extenuating Circumstances procedure](https://myport.port.ac.uk/my-course/extenuating-circumstances) is there to support you if you have had any circumstances (problems) that have been serious or significant enough to prevent you from attending, completing or submitting an assessment on time. If you complete an Extenuating Circumstances Form (ECF) for this assessment, it is important that you use the correct module code, item number and deadline (not the late deadline) given above.
* [ASDAC](http://www2.port.ac.uk/additional-support-and-disability-advice-centre/) are available to any students who disclose a disability or require additional support for their academic studies with a good set of resources on the [ASDAC moodle site](https://moodle.port.ac.uk/course/view.php?id=3012)
* The University takes any form of academic misconduct (such as plagiarism or cheating) seriously, so please make sure your work is your own. Please ensure you adhere to our [*Student Conduct Policy*](https://policies.docstore.port.ac.uk/policy-261.pdf) and watch the video on [Plagiarism.](https://www.youtube.com/watch?v=2a0QJnCmfEs)
* Any material included in your coursework should be fully cited and referenced in **APA 7** format. Detailed advice on referencing is available from the [library](https://library.port.ac.uk/w165.html), also see [TECFAC 08 Plagiarism](https://www.youtube.com/watch?v=2a0QJnCmfEs).
* Any material submitted that does not meet format or submission guidelines, or falls outside of the submission deadline could be subject to a cap on your overall result or disqualification entirely.
* If you need additional assistance, you can ask your personal tutor, student engagement officer [ana.baker@port.ac.uk](mailto:ana.baker@port.ac.uk), academic tutor [eleni.noussi@port.ac.uk](mailto:eleni.noussi@port.ac.uk) or your lecturers.
* If you are concerned about your mental well-being, please contact our [Well-being service](https://myport.port.ac.uk/guidance-and-support/health-and-wellbeing)

## **Presentation instructions**

The presentation will be 10-12 minutes long, with 3-5 minutes questions.

You are allowed to use any software to create your presentation.

You **must** use the Tior Games database to do this coursework.

Tior Games Case study

Tior Games is dedicated to providing an exceptional player experience as its primary objective. To ensure a consistent, fast, and always-available high-quality player experience for its global user base, Tior Games focuses on its most popular and lucrative game, League of Fun (LoF). LoF, a free-to-play game with in-game purchases, operates as a computer-based MOBA.

Given the widespread popularity of LoF, Tior Games hosts World Championships biannually, in January and September, for competitive and professional teams. The awards and event popularity have grown each year, mirroring the overall rise in gaming industry and community interest.

A few years ago, Tior Gaming migrated its entire data warehouse to Amazon AWS, specifically using Hive. As a user with access to a segment of the company's data warehouse managed by their DBA, the data is centered around the previous World Championships starting from 2016. The structure employs a snowflake schema, encompassing game data, champion data, player information, and event details.

The objective is to analyze the accessible data and propose strategies for Tior Gaming to enhance both player and spectator experiences, boost the World Championship's popularity, and overall improve profitability. The proposed solution involves querying the data within the data warehouse and subsequently creating a visual business presentation for Tior's Board of Executives.

**To access the data warehouse:**

Connect to the university’s VPN.

Go to Azure data studio or VSCode, connect with your login details as we did in the workshop and are detailed on Moodle.

**Select the TiorGames database. You are NOT allowed to use any other data warehouse.**

Keep in mind that the data warehouse has issues as it happens in real life, and this is to be expected. You are unable to edit or “clean” the data given to you.

The data dictionary is also provided in a separate file link:

<https://docs.google.com/spreadsheets/d/1-ifRuypTcBd_NZAR_fBydqIvi7IiVEbp/edit?usp=sharing&ouid=100867728873298265982&rtpof=true&sd=true>

# **Data warehouse**

**You MUST use the Tior Games data warehouse.**

**You will be asked to run the code produced, so please have all the code ready to run to avoid delays.**

**Before the presentation and before the deadline: Create and submit on the dropbox in Moodle one SHORT slide presentation with all your answers.**

During the presentation:

* You will execute all your code and discuss the rationale behind it.
* You will have to be able to discuss the answers to the questions.
* You are not allowed to have any notes other than your report.
* The presentation will be 5-8 minutes in length plus 2 minutes for questions and it will be on a 1-1 basis.

**You must answer all of the following questions:**

## **Question 1**

Compose and execute **six** T-SQL queries. In the slide presentation, include the code for each query.

*Notes*:

* Ensure that your queries utilize OLAP functions and incorporate data warehouse-specific keywords, as discussed in the workshops. OLTP queries will receive zero marks.
* Make your queries meaningful, showcasing the capability of a Data Warehouse in supporting decision-makers.
* Employ a diverse range of fact tables and dimensions from the provided Data Warehouse.
* Provide brief explanation of the business rationale behind its query.
* Each of the six queries must differ from one another, utilizing distinct fact tables and dimensions.
* Ensure that each query incorporates at least one data warehouse concept. Queries resembling "select \* from tablename where condition is true" and repetitive queries will be scored as zero.
* Avoid presenting a query that answers question 3 in this assignment.

{Total marks: 48}

## **Question 2**

Revise the given schema and propose a minimum of two additional dimensions that would yield valuable insights.

In the slide presentation, include the two dimensions, their data dictionaries, and a single sentence justifying their inclusion.

*Notes*:

* Articulate the rationale and business reasoning behind your choices, explaining why these dimensions and how the business can leverage insights derived from them.
* Present these two dimensions using the existing naming convention in the data warehouse.
* Ensure these two dimensions seamlessly integrate with at least one fact table or dimension in the current data warehouse.
* Include a data dictionary following the same format as the provided data dictionary.
* Utilize any visual means you prefer to present the dimensions and their data dictionaries.

{Total marks 20}

## **Question 3**

Utilize the provided data warehouse to formulate ONE T-SQL query answering the following questions:

a. Determine the online item that experienced the most unexpected sales in Japan and identify the corresponding year.

b. Identify the online item with the least unexpected sales when considering all countries and years. Specify the country and exact year for this occurrence.

*Notes*:

* Think: what might be considered unexpected in the sales business.
* Assume the existence of required indexes and constraints for the tables.
* Craft your queries with clarity and efficiency.

In the slide presentation include the code of the query and the 2 answers.

{Total marks 12}

## **Question 4**

External dataset integration.

Based on the data warehouse provided. You are tasked to identify and describe the rationale of an external dataset or identify a real company that their data would be able to complement Tior Games data warehouse.

*Instructions*:

1. Identify an external dataset or a real company with relevant data that, when integrated into Tior Games' data warehouse, could offer valuable insights.
2. Provide details on the nature of the external dataset or the identified company, including the type of data they collect, its scope, and its potential relevance to Tior Games.
3. Clearly articulate the rationale behind selecting this external dataset or company.
4. Describe the specific benefits that integrating this external data would bring to Tior Games.
5. Anticipate potential challenges or issues that may arise during the integration process. Propose mitigation strategies or solutions to address these challenges and ensure a smooth integration.
6. Summarize your findings in a presentation format suitable for Tior Games' Board of Executives. Include key visuals, and referneces to illustrate the potential impact of integrating the external dataset on Tior Games' data warehouse.

{Total marks 20}

**Marking scheme**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Marks** | **Task** | **Fail (<40%)** | **Pass (40-49%)** | **Pass (50-59%)** | **Good Pass (60-69%)** | **Excellent (70% +)** |
| **Criteria for all tasks** | | Inadequate English, inadequate presentation, no slides presented or completely wrong or only minimal understanding. Does not follow submission instructions. |  | Mostly correct presentation with mostly correct UK English grammar and spelling on slides. Complies with most submission instructions. | Correct UK English grammar and spelling with clear writing style and good presentation. Complies with all submission instructions. | Correct UK English grammar and spelling. Excellent structure and presentation. Complies with all submission instructions |
| **Question 1**  **[Total 48 marks]** | | | | | | |
| **Per query**  **[8 marks per query, criteria below]** | | | | | | |
| **3** | **Incorporation of Data Warehouse Concepts (OLAP)** | Fails to incorporate relevant data warehouse concepts | Suboptimal use of resources. Limited or unclear incorporation of concepts | Incorporates some data warehouse concepts | Effectively incorporates at least one concept | Incorporates multiple data warehouse concepts effectively |
| **2** | **Meaningfulness & Business Insight** | Little to no meaningful insights provided | Limited meaningful insights, needs improvement | Adequate insights but may lack depth | Insights are meaningful but could be more impactful | Query provides meaningful insights for decision-making |
| **2** | **Business Rationale Explanation** | Poor or missing explanation of the business rationale | Explanation lacks clarity or depth | Offers an adequate rationale with some room for improvement | Provides a clear explanation of the rationale | Clearly explains the business rationale behind the query |
| **1** | **Overall, avoidance of Repetitive Queries & diversity of dimensions and fact tables & query Structure & clarity** | Submits queries with high repetition and lack of variety.  Poorly structured and difficult to follow | Significant overlap or repetitive queries. And/or utilizes limited fact tables and dimensions. Structure lacks clarity and organization | Some overlap between queries and dimensions and fact tables, but generally distinct. Adequate structure but some room for improvement | Provides distinct queries with distinct dimensions and fact tables with minor overlap. Clear and well-structured query. | Ensures that queries and dimensions are distinct and not repetitive. Exceptionally clear and well-structured query |
| **Exercise 2**  **[Total 20 marks]** | | | | | | |
| **2** | **Additional Dimensions** | Student fails to propose relevant dimensions | Student proposes one relevant dimension | Student proposes two dimensions with some relevance | Student proposes two relevant dimensions | Student proposes two highly relevant dimensions |
| **8** | **Rationale & Business Justification** | Inadequate or missing justification | Limited or unclear justification | Adequate justification with some gaps | Clear and compelling justification | Exceptionally clear and thorough justification |
| **2** | **Consistent Naming Convention** | Complete disregard for naming convention | Significant inconsistencies in naming convention | Some inconsistencies in naming convention | Generally consistent with minor deviations | Consistent use of existing naming convention |
| **4** | **Integration with Current Data Warehouse** | No attempt at integration provided | Poor integration with existing structures | Integration has some challenges | Generally, integrates well, with minor issues | Seamless integration with existing fact table/dimension |
| **2** | **Data Dictionary Format & Completeness** | Data dictionaries are incomplete or absent | Limited data dictionaries, missing key elements | Adequate data dictionaries, with some gaps | Data dictionaries are mostly complete and clear | Data dictionaries are comprehensive and well-structured |
| **2** | **Overall Impressions & Creativity** | Lack of creativity and minimal effort | Limited creativity or missing elements | Adequate creativity in dimension choices and presentation | Demonstrates creativity in dimension choices and presentation | Outstanding creativity in dimension choices and presentation |
| **Question 3**  **[Total 12 marks]** | | | | | | |
| **3** | **Query Structure & Clarity** | Poorly structured and difficult to follow | Structure lacks clarity and organization | Adequate structure but some room for improvement | Clear and well-structured query | Exceptionally clear and well-structured query |
| **6** | **Accuracy of Results** | Results are unreliable and highly inaccurate | Significant inaccuracies impacting result analysis | Some inaccuracies, affecting result interpretation | Mostly accurate with minor discrepancies | Results are accurate and align with business expectations |
| **2** | **Inclusion of Country and Year Information** | Missing or inadequate information | Limited or unclear inclusion of country and year details | Adequate inclusion but could be presented more clearly | Includes necessary information but may lack clarity | Clearly includes country and year information in the output |
| **1** | **Slide Presentation - Code Presentation** | Code presentation is missing or incomplete | Code is presented poorly, difficult to comprehend | Adequate code presentation with room for improvement | Code presentation is good but may lack some clarity | Code is well-presented, clear, and easy to follow |
| **Question 4**  **[Total 20 marks]** | | | | | | |
| **2** | **Identification of External Dataset/Company** | Fails to identify a relevant dataset or company | Identifies a dataset or company with limited relevance | Identifies a dataset or company with some relevance | Identifies a relevant external dataset or company and has clear data scope relevance. | Identifies an external dataset or company with highly relevant and comprehensive data. |
| **6** | **Rationale for Selection** | Poor or missing rationale | Limited or unclear rationale | Adequate rationale with some room for improvement | Clear and convincing rationale | Exceptionally clear and compelling rationale |
| **6** | **Benefits of Integration** | Fails to describe relevant benefits | Describes limited or unclear benefits | Describes benefits with some clarity | Describes benefits with clarity and detail | Clearly articulates specific benefits for Tior Games |
| **4** | **Anticipation of Challenges** | Fails to anticipate challenges or propose solutions | Anticipates challenges with limited or ineffective solutions | Anticipates challenges with some proposed solutions | Anticipates challenges and proposes reasonable solutions | Anticipates potential challenges and proposes effective mitigation strategies |
| **2** | **Presentation Quality & Creativity** | Poor presentation lacking key elements | Limited or unclear presentation format | Adequate presentation with some room for improvement | Good presentation format with clear visuals | Exceptional presentation format suitable for executives, including key visuals and references |