

BDMM - 2nd Project(2)

May 6, 2021

1 Big Data Modeling and Management Assignment

1.1 MongoDB first homework

During the second and third homeworks we will be looking into a staple of SQL databases the **World Wide Imports** database create by Microsoft to experiment with MSSQL.

For this first project we will be focusing on querying and analysing data on 3 collections **orders**, **customers** and **users**.

In short: *Wide World Importers (WWI) is a wholesale novelty goods importer and distributor operating from the San Francisco bay area.*

Further detail and information on this database can be found here: (<https://docs.microsoft.com/en-us/sql/samples/wide-world-importers-what-is?view=sql-server-ver15>)

Problem description Understand the data.

Answer the questions.

Submit the results by following the instructions

Connection details to the mongodb database

Host: `rhea.isegi.unl.pt:27017`

Database: `worldwideimporters`

Username: `{groups_username}`

Password: `{groups_password}`

Connection URI: `mongodb://{groups_username}:{groups_password}@rhea.isegi.unl.pt:27017/worldwideimporters`

Credentials sent by email.

Questions

0. **Example Question** *How many orders exist in the database?*
1. How many people records don't have the `UserPreferences` field?
2. How many customer records are valid after `November 2015`?
3. How many people have their `Title` equal to `Team Member`?
4. How many people have in their name the string `Sara`?
5. Return 5 full names that have in their name the string `Sara`?
6. What is the highest `CommissionRate` that a person has?
7. And what are the top 10 most Common Names (Primary or Surnames)?
8. How many orders has the Customer `Tailspin Toys (Head Office)`?

9. How many people that have more or equal than three `OtherLanguage`?
10. Top 10 most common `OtherLanguage` for people records?
11. Who is the most common `PickedByPersonID` person name for orders done by customer `Adriana Pena`?
12. What is the average difference in days between `OrderDate` and `ExpectedDeliveryDate` for orders sold by (`SalespersonPersonID`) person with name `Jack Potter`?

Groups Groups should have 4 to 5 people

You should register your group on moodle. An email will be going out to everyone with the credentials for the database to use when storing the results.

Submission Upload the notebook with the results in moodle before **23:59 of May 16nd**

Evaluation This will be 20% of the final grade.

Each solution will be evaluated on 2 components: correctness of results and simplicity of the solution.

All code will go through plagiarism automated checks. Groups with the same code will undergo investigation.

Note: Remember the MongoDB is a shared database, use limits when running your queries.