Assignment

Task 1

The file 2014-09-03 log is a log file from a demo environment, including all entries for a particular day.

Using this log as input, please answer the following questions.
1. List of URLs that were not found (404 error), including number of times each URL was requested
2. Average Server Response Time
3. Which database table is most frequently loaded?
4. Is any URL redirection taking place?
5. Are there any server errors? Ideas about possible causes?

Deliverables
For each question, please provide the following:
- Answer
- Explanation of how you arrived at the result, including any scripts or other commands that you used.

Task 2

You are working for a business that owns two DVD rental stores. Both stores use software to manage the DVD rentals, but the software has no reporting capabilities.
The owner wants to better understand various aspects of his business, and has asked you a few questions.
In order to answer his questions, you will design and execute a few queries on the database of the DVD rental software that is used by the stores (there is one database for both stores). You can view the database schema at the following link:
http://www.postgresqltutorial.com/postgresql-sample-database/
This page also contains instructions to download the full database and restore it in a PostgreSQL instance that you will set up.

Please provide the SQL queries that you will use to answer the owner's questions, as well as the actual answers.

1. Which customer has made the most rentals at store 2?
2. A customer tried to rent "Image Princess" from store 1 on 29/07/2005 at 3pm but it was sold-out. Would he be able to rent it from store 2 if he had tried?
3. How many customers are active at any given month per year (e.g. …, Jun 2005, Jul 2005,...., Jun 2006 etc)? We define active as performing at least one rental during that month.
4. Calculate the average length of films in the database and List the films and their lengths where the length is greater than the average length of all films
5. List the customers who have made more payments than the average number of payments.
6. Are there any other insights that you can gather from the data that would be helpful to the owner of this business?

Hints
You may receive a few errors on alter table statements when restoring the database, which can be ignored. You may also find it useful to use the WITH statement.

Deliverables
For each question, please provide the following:
- Answer
- SQL query
- For question 6, also explain the way in which the insights that you provide will be useful to the business

---

Task 3
Your objective is to build a small project named MovieData. The application will use the API from The Movie DB (https://developers.themoviedb.org/3/getting-started). Use the following API key: bbb0e77b94b09193e6f32d5fac7a3b9c or create a new one on the site.

The goal of this assignment is to implement an application or script that will retrieve and store movie information.

The application must retrieve a list of the movies currently in theaters in Greece, along with at least the following attributes:
- **Title**
- **Description**
- **Original Title**
- **List of Directors**
- **The IMDB link to the profile of each of the directors**

All the above information will be stored in a relational database. The data in the database will be updated each time the process runs.

You are free to include any additional features / optimizations that you may find relevant or could showcase your skills but please bear in mind that you should cover the core requirements first before attempting any improvements. Also note that, although it may be tempting to use ready-made libraries or gems for querying TMDB, we would prefer you to make direct API requests.

Deliverables

The final deliverable should contain:
- Source code.
- DDL for Database schema.
- A simple Readme.txt that will describe the way to build and use the application along with any other external systems that you may have used.

You will be assessed on the following:
a) Fitness for purpose (i.e. your application does what its specifications require)
b) Simplicity (the simpler / smaller the solution, the better)
c) Code quality