The requirements for the project, then, are as follows:

**1. A command-line client application appropriate to the domain**

Requirements:

1. Querying the data in a way that a customer in the domain would do

2. Modifying the data in a way that a customer in the domain would do

Decide:

1. What you think an ideal client should be able to do

2. What you plan to actually implement for your client given the time constraints

3. (At the end of the project, when you write your report) What you actually

implemented from your plan, and what you left

4. An explanation justifying each of the above choices

**2. An entity-relationship design to model the data**

1. All entity sets, specifying the entity set name and attributes, showing any

compound attributes, multivalued attributes, and optional attributes per the

methods described in the course

2. All relationship sets, specifying the relationship set name and any attributes

it might have

3. All primary keys, cardinality constraints, and attribute domains

4. Any weak, specialized, or aggregations

5. Any other aspects relevant to an ER design

**3. A relational schema based on the ER design**

* write the necessary SQL code to:

1. create the required tables, views, etc. for the relational schema

2. create the required primary keys, foreign keys, and integrity constraints

3. create indexes as necessary for the query operations you will do both in the

client and in the data-mining exercise

4. load the data from your dataset CSVs into the tables

* Handle error and errors and inconsistencies relative to your design in the data:

1. fixing obvious data errors

2. removing any duplicate data

3. modifying your design in certain cases

Note: remember what you did to handle the issues

**4. A data-mining investigation of the dataset**

1. Select a domain-appropriate question that you want data mining to answer

2. Select a technique or technique that will be appropriate to the question you

are investigating

3. Implement said technique efficiently to build a data model

4. Determine the validity of your model

5. Report the results of your investigation

**Deliverables:**

**1. Final Written Report:** this should describe the client application, the ER

design, the relational schema, and your data-mining investigation, detailing

the specific issues required above. In addition, you should include a test case

plan that describes how you test the various code aspects of your project.

**2. Code:** you are not expected to submit your code but rather store it in a

GitHub repository, or equivalent, to which we can have access. The code in

the repository should include the following:

(a) All client code

(b) The SQL code necessary to implement the relational schema and load

the data from the CSV files

(c) The code, SQL and otherwise, needed to implement your data-mining

investigation

(d) Test cases for the above

**3. Video Demo:** a 20-minute walk-through/presentation of your project. It

should describe all of the aspects of your design, implementation, and results.