

Project Appendix

Group 2 – Hao Niu, Ke Wang, Tianyu Fang, Yao Cheng

Introduction

This is a project resembles a real-world System-Analysis Life-Cycle process, where a database - regarding bird-strike incidents - is designed, implemented, and then analyzed to provide insights and responses to a set of queries, as requested, that are relevant to the aviation industry.

The project primarily involves the following:

1. Conceptual & Logical modeling

Both a UML class diagram and an ER diagram were designed to reflect the overall structure of the database. The UML class diagram was used to represent the conceptual model, where all the necessary entities, attributes, relationships, along with multiplicities were included. The ER diagram was used to represent the logical model, and except what was provided in the UML class diagram, it also included the primary and foreign keys, and the data types of the attributes.

2. Data realization & implementation

The above mentioned conceptual and logical models were then implemented into a SQLite database, via the use of Google Colab Notebook, with a mix of SQLite and Python code.

3. Data loading & cleaning

The database was programmatically populated from a CSV file provided, all the information contained in the CSV file was firstly reviewed and cleaned, before being loaded into the database. For the detailed data selection process, please refer to the next section.

4. Data querying & result analysis

In the last two parts of the project, a set of queries were formulated and executed on the database, to firstly retrieve the necessary information as requested, and to offer further insights based on analyses from other perspectives.

Data selection rationale

This section provides a detailed explanation for the selection of the data that was used in the project. The original dataset was provided in a CSV file, which contained a total of 26 columns and over 25,000 rows, about 300 of which were deemed invalid due to missing or incorrect data, and 12 columns were considered irrelevant for the purpose of this project. The remaining information was then loaded into 7 tables namely: Birds, WeatherCondition, StrikeRecord, Flight, Airport, AircraftModel and Airline.

The overall selection process was based on the following criteria:

1. Do the selected columns contain relevant and sufficient information for the required scenario below?

"Your team's fictional client is America's Federal Aviation Authority (FAA), who needs a database built to help them organize and analyze the data they've collected on bird-strike incidents ...

For each incident, we need to know when it happened, how many birds were involved, and during which flight phase and at what altitude the incident happened. We need to know what airport was involved and what state the airport is in. We need to know details about the plane as well as the airline which operates the plane. We need to know the weather and precipitation going on during the incident. We need to know some details about the kind of birds involved in the incident."

2. Will the selected columns help in answering the following questions as requested?
 - Find the number of bird strike incidents for each airline upon take-off or climb. Include all airlines.
 - Find which airport had the most birds involved in incidents. The total number of birds, not incidents, is what matters here. Can be during any flight phase.
 - Find the number of bird strike incidents by year. Include all airlines.
3. Could the selected columns be used to provide additional insights or analyses, like "the number of bird strikes incidents per year from 2008 to 2011 during take-off/climbing and during descent/approach/landing"?
4. Besides the above, while maintaining data integrity, could the original 26 * 25,000 dataset be reduced to a more manageable size, without losing the essence of the data?

In general, all the previously mentioned criteria were taken into consideration, and the final selection of columns was made to ensure that the data was relevant, sufficient, and could be used to answer the queries as requested, while also providing additional insights and analyses which are beyond the scope of the requested queries.

Conclusion

To conclude, this project on the bird-strike incidents database was designed, implemented, and analyzed to provide insights and responses to a set of queries, as requested, that are relevant to the aviation industry. The project involved the conceptual and logical modeling of the database, the implementation of the models into a SQLite database, the loading of data from a CSV file, and the querying of the database to retrieve the necessary information. The project also provided additional insights and analyses based on the data, which were beyond the scope of the requested queries. The project was successful in achieving its objectives and in providing valuable information to the fictional client, America's Federal Aviation Authority (FAA).