Problem statement taken

Load the Wikidata into ArangoDB and create a graph that represents the relationships between entities, such as people, places, and things. Use AQL to query the graph and find all instances of a particular entity type, such as all instances of a person.

Introduction

Over the past few years, the hotel sector has experienced rapid growth, which has increased consumer demand for tailored recommendations. We have created a hotel recommendation system based on customer preferences in this data science assignment adding a railway dataset to see the connectivity of railways along with it.

Methodology

In order to find trends in user behavior and preferences, our strategy entails examining big datasets of hotel reviews and ratings, close to railway and tourism locations. Then, utilizing machine learning algorithms, we create a model that can precisely forecast which hotel guests will likely enjoy.

To do so we first gather a sizable dataset of hotels, railroads, and tourist destinations in order to construct our recommendation engine. This information will be collected from well-known travel websites like TripAdvisor and Booking.com using a secondary data platform.

After the data has been gathered, it is cleaned and preprocessed to weed out any extraneous information and guarantee consistency across the dataset. Removing duplicates, standardizing rating scales, and collecting pertinent details such as hotel facilities and location are among the duties involved in this.

Technical Details of our Database:
ArangoDB is a multi-model database that supports document, key-value, and graph data models. It provides a powerful graph database functionality that allows users to store, manage, and analyze graph data efficiently. Here are some ways in which Our team has tried to visualize data and analyze data using ArangoDB:

- **Graph data storage:**
  ArangoDB provides a native graph data model that allows users to store graph data in a highly efficient manner. It supports both directed and undirected graphs and allows users to define custom properties for vertices and edges.

- **Graph traversal:**
  ArangoDB provides a powerful graph traversal API that allows users to traverse graphs using various algorithms such as depth-first search, breadth-first search. This API can be used to perform complex graph analysis tasks such as finding the shortest path between two vertices or identifying clusters within a graph.

- **Graph visualization:**
  ArangoDB provides built-in support for graph visualization. This allows users to create interactive visualizations of their graph data, making it easier to understand and analyze.

- **Arango Search:**
  It is a feature of ArangoDB that allows users to perform full-text search and ranking on data stored in ArangoDB. It is built on top of ArangoDB's native indexing capabilities and provides a powerful and flexible search engine that can be used to search across multiple collections and data types

- **Connecting with Python:**

Connecting Arangodb with python(jupyter notebook using pyArango library, plotting geo locations using Folium library and the graphs on jupyter notebook using networkx library of python

Results and Discussion

We experimented with various machine learning algorithms, including decision trees, random forests, and K nearest neighbors. We evaluated the models using performance tools such as accuracy score, best performance and precisions .

Conclusion

- In this data science project, we aim to develop a personalized hotel recommendation system based on user preferences. Through data collection, exploratory data analysis, machine learning model development, and user interface design, we hope to create a system that is accurate, relevant, and easy to use

- According to Feature importance, the percentage of people preferring hotels on the basis of hotel room is **11.08**, on the basis of nearby places is **7.29**, on the basis of site review rating is **6.99**.

- By **EDA**, the maximum number of customers prefer hotels near the railway station and the market, but do not prefer hotels near the airport.

- The data was loaded into ArangoDB, and a graph that represents the relationship between entities such as people, places and things. The graph was queried with AQL

- This recommendation system streamlines the manual hotel search process, saves time, and enables the user to choose more effectively based on their preferences.

References

- Getting Started with ArangoDB (Free Course on Udemy)
- Study guide for Udemy course
- Getting Started with Graphs in ArangoDB (Free Graph Course)
- ArangoDB Interactive Tutorials (Python Notebooks)
- Lunch & Learn videos
- ArangoML (Graph Powered Machine Learning)
  1. ArangoML - Multi-Model Collaboration
  2. ArangoDB for Graph ML
  3. Graph Embeddings Tutorial (Python notebook)
  4. Introducing the ArangoDB - DGL Adapter
  5. Introducing the Arango NetworkX Adapter (NetworkX example)