ETL Project Proposal – TrueSafeCars.com

Team NoFloods (Group 8)
➢ Roopa Patel
➢ Samuel Parks
➢ Steven Lee

Overview

We are a new nationwide used car dealership, **TrueSafeCars.com**, that is focused on selling the safest used cars available. We were formed by buying out two major used car dealership chains that were in financial trouble. Our goal is to provide our customers with safe, quality cars that are less than 20 years old.

With every purchase, we provide buyers with a list of manufacturer recalls specific to their vehicle. Our Service department works with every buyer to make sure these recalls are addressed.

To meet our company's goals, we need an ETL solution to perform the following:
1. Consolidate the inventory into a central database
2. Assign the inventory to our 5 main offices
3. Identify which vehicles meet our safety guidelines
4. Identify recall information per vehicle
5. Identify which vehicles should be sold at auction (vs. selling to the public)

Extract Requirements

Pull vehicle inventory data from the two dealerships we purchased:
- Data for purchased dealership #1: Kaggle.com (https://www.kaggle.com/austinreese/craigslist-carstrucks-data)
- Data for purchased dealership #2: Kaggle.com (https://www.kaggle.com/jpayne/852k-used-car-listings)

Transform Requirements

Once all of the vehicle inventory data has been extracted, perform the following transformations:
- Remove duplicate VIN numbers
- Assign the inventory to our 5 main office locations. We will provide a reference table to provides the State <-> Main Office mapping. Our 5 main office locations are:
  o Northeast Main Office
  o Southeast Main Office
  o Midwest Main Office
  o Southwest Main Office
  o West Main Office
- Assign each vehicle an overall safety rating using an API from NHTSA.gov or similar
  o https://webapi.nhtsa.gov/Default.aspx?SafetyRatings/API/5
- Determine the number of open recalls per vehicle using an API from NHTSA.gov or similar
  o https://webapi.nhtsa.gov/Default.aspx?Recalls/API/83

The vehicle inventory should have a flag/indicator to show if the vehicle is ready to be sold to the public or not:
- Vehicles at 20 years old or older should be flagged for auction
- Vehicles with a poor safety rating should be flagged for auction
- Vehicles in poor condition (if this information is available in the inventory data) should be flagged for auction

Load Requirements

Load transformed data into a central database (NoSQL or SQL is to be determined). The central database should have tables that can be easily queried to retrieve:
- Vehicles assigned to each main office
- Vehicles that are ready to be sold to the public
- Vehicles that should be sold at auction
- Overall safety rating for each vehicle
- Recalls that apply to each vehicle