**Project Management Document: Power BI Dataset Preparation**

**Overview**

This document details the preparation and structure of several datasets designed for use in Power BI. The datasets are organized into a star schema, including a central fact table and multiple dimension tables. These datasets will facilitate in-depth analysis of trip data, vehicle usage, driver performance, and the impact of fuel prices.

**Datasets**

**1. Date Dimension Table (date\_dim.csv)**

* **Description**: This table contains date-related attributes for time-based analysis.
* **Columns**:
  + Date: The specific date (datetime format).
  + Year: The year of the date.
  + Month: The name of the month.
  + Weekday: The name of the day of the week.
  + WeekOfYear: The week number within the year.
  + Quarter: The quarter of the year (1-4).

**2. Driver Dimension Table (driver\_dim.csv)**

* **Description**: Contains driver information linked to their respective cardID.
* **Columns**:
  + cardID: Unique identifier for the driver.
  + Driver\_Name: Name of the driver (randomly generated for simulation).
  + Business\_Unit: The business unit assigned to the driver (A, B, or C).

**3. Vehicle Dimension Table (vehicle\_dim.csv)**

* **Description**: Contains details about the vehicles used in trips.
* **Columns**:
  + Unit\_Name: The identifier for the vehicle unit.
  + VIN: Vehicle Identification Number.
  + Car\_Model: The model of the car.
  + Car\_Brand: The brand of the car.

**4. Fact Trips Table (fact\_trips.csv)**

* **Description**: The central fact table containing detailed information about each trip.
* **Columns**:
  + №: Unique identifier for each trip.
  + NBLbutton1: Operational code linked to the trip.
  + cardID: Unique identifier for the driver.
  + Beginning: The start date and time of the trip.
  + End: The end date and time of the trip.
  + Mileage: Distance covered during the trip (in km).
  + Duration: Duration of the trip (timedelta format).
  + Driver: Placeholder data field (anonymized).
  + Urban mileage: Distance covered in urban areas (in km).
  + Suburban mileage: Distance covered in suburban areas (in km).
  + VIN: Vehicle Identification Number (linked to the vehicle\_dim table).
  + Weekday: Day of the week when the trip began.
  + WeekOfYear: Week number in the year when the trip began.
  + Trip Length Category: Categorized as Short, Medium, or Long based on mileage.
  + Trip Duration Category: Categorized as Short, Medium, or Long based on duration.
  + Driver\_Name: Name of the driver (linked to the driver\_dim table).
  + Business\_Unit: The business unit of the driver (linked to the driver\_dim table).

**5. Fuel Prices Dimension Table (fuel\_prices\_dim.csv)**

* **Description**: Contains daily fuel price information, linked to the date dimension.
* **Columns**:
  + Date: The specific date (datetime format).
  + 95-Octane Gasoline (SEK/liter): Price of 95-octane gasoline in SEK per liter.
  + Diesel (SEK/liter): Price of diesel in SEK per liter.
  + Fast Charging EV (SEK/kWh): Price for fast charging electric vehicles in SEK per kWh.

**Usage and Relationships**

* **Star Schema Structure**: The fact\_trips table is at the center, connected to dimension tables (date\_dim, driver\_dim, vehicle\_dim, fuel\_prices\_dim).
* **Relationships**:
  + fact\_trips <-> driver\_dim via cardID.
  + fact\_trips <-> vehicle\_dim via VIN.
  + fact\_trips <-> date\_dim via Beginning (linked to Date).
  + fact\_trips <-> fuel\_prices\_dim via Beginning (linked to Date).