**Database Design**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| ID | INT | Primary Key, unique identifier for each record |
| Brnd\_Name | VARCHAR(255) | Brand name of the drug |
| Gnrc\_Name | VARCHAR(255) | Generic name of the drug |
| Tot\_Mftr | INT | Total number of manufacturers |
| Mftr\_Name | VARCHAR(255) | Name of the manufacturer |
| Tot\_Spndng\_2018 | DECIMAL(15,2) | Total spending in 2018 |
| Tot\_Dsg\_Unts\_2018 | DECIMAL(15,2) | Total dosage units in 2018 |
| Tot\_Clms\_2018 | INT | Total claims in 2018 |
| Tot\_Benes\_2018 | INT | Total beneficiaries in 2018 |
| Avg\_Spnd\_Per\_Dsg\_Unt\_Wghtd\_2018 | DECIMAL(15,2) | Average spending per dosage unit (weighted) in 2018 |
| Avg\_Spnd\_Per\_Clm\_2018 | DECIMAL(15,2) | Average spending per claim in 2018 |
| Avg\_Spnd\_Per\_Bene\_2018 | DECIMAL(15,2) | Average spending per beneficiary in 2018 |
| Outlier\_Flag\_2018 | BOOLEAN | Indicates if the data is an outlier in 2018 |
| Tot\_Spndng\_2019 | DECIMAL(15,2) | Total spending in 2019 |
| Tot\_Dsg\_Unts\_2019 | DECIMAL(15,2) | Total dosage units in 2019 |
| Tot\_Clms\_2019 | INT | Total claims in 2019 |
| Tot\_Benes\_2019 | INT | Total beneficiaries in 2019 |
| Avg\_Spnd\_Per\_Dsg\_Unt\_Wghtd\_2019 | DECIMAL(15,2) | Average spending per dosage unit (weighted) in 2019 |
| Avg\_Spnd\_Per\_Clm\_2019 | DECIMAL(15,2) | Average spending per claim in 2019 |
| Avg\_Spnd\_Per\_Bene\_2019 | DECIMAL(15,2) | Average spending per beneficiary in 2019 |
| Outlier\_Flag\_2019 | BOOLEAN | Indicates if the data is an outlier in 2019 |
| Tot\_Spndng\_2020 | DECIMAL(15,2) | Total spending in 2020 |
| Tot\_Dsg\_Unts\_2020 | DECIMAL(15,2) | Total dosage units in 2020 |
| Tot\_Clms\_2020 | INT | Total claims in 2020 |
| Tot\_Benes\_2020 | INT | Total beneficiaries in 2020 |
| Avg\_Spnd\_Per\_Dsg\_Unt\_Wghtd\_2020 | DECIMAL(15,2) | Average spending per dosage unit (weighted) in 2020 |
| Avg\_Spnd\_Per\_Clm\_2020 | DECIMAL(15,2) | Average spending per claim in 2020 |
| Avg\_Spnd\_Per\_Bene\_2020 | DECIMAL(15,2) | Average spending per beneficiary in 2020 |
| Outlier\_Flag\_2020 | BOOLEAN | Indicates if the data is an outlier in 2020 |
| Tot\_Spndng\_2021 | DECIMAL(15,2) | Total spending in 2021 |
| Tot\_Dsg\_Unts\_2021 | DECIMAL(15,2) | Total dosage units in 2021 |
| Tot\_Clms\_2021 | INT | Total claims in 2021 |
| Tot\_Benes\_2021 | INT | Total beneficiaries in 2021 |
| Avg\_Spnd\_Per\_Dsg\_Unt\_Wghtd\_2021 | DECIMAL(15,2) | Average spending per dosage unit (weighted) in 2021 |
| Avg\_Spnd\_Per\_Clm\_2021 | DECIMAL(15,2) | Average spending per claim in 2021 |
| Avg\_Spnd\_Per\_Bene\_2021 | DECIMAL(15,2) | Average spending per beneficiary in 2021 |
| Outlier\_Flag\_2021 | BOOLEAN | Indicates if the data is an outlier in 2021 |
| Tot\_Spndng\_2022 | DECIMAL(15,2) | Total spending in 2022 |
| Tot\_Dsg\_Unts\_2022 | DECIMAL(15,2) | Total dosage units in 2022 |
| Tot\_Clms\_2022 | INT | Total claims in 2022 |
| Tot\_Benes\_2022 | INT | Total beneficiaries in 2022 |
| Avg\_Spnd\_Per\_Dsg\_Unt\_Wghtd\_2022 | DECIMAL(15,2) | Average spending per dosage unit (weighted) in 2022 |
| Avg\_Spnd\_Per\_Clm\_2022 | DECIMAL(15,2) | Average spending per claim in 2022 |
| Avg\_Spnd\_Per\_Bene\_2022 | DECIMAL(15,2) | Average spending per beneficiary in 2022 |
| Outlier\_Flag\_2022 | BOOLEAN | Indicates if the data is an outlier in 2022 |
| Chg\_Avg\_Spnd\_Per\_Dsg\_Unt\_21\_22 | DECIMAL(10,4) | Change in average spending per dosage unit from 2021 to 2022 |
| CAGR\_Avg\_Spnd\_Per\_Dsg\_Unt\_18\_22 | DECIMAL(10,4) | Compound Annual Growth Rate of average spending per dosage unit from 2018 to 2022 |

**Explanation**

* **ID**: A unique identifier for each record to act as the primary key.
* **Brnd\_Name**: The brand name of the drug.
* **Gnrc\_Name**: The generic name of the drug.
* **Tot\_Mftr**: The total number of manufacturers for the drug.
* **Mftr\_Name**: The name of the manufacturer.
* **Tot\_Spndng\_20XX**: Total spending for the year specified.
* **Tot\_Dsg\_Unts\_20XX**: Total dosage units for the year specified.
* **Tot\_Clms\_20XX**: Total claims for the year specified.
* **Tot\_Benes\_20XX**: Total beneficiaries for the year specified.
* **Avg\_Spnd\_Per\_Dsg\_Unt\_Wghtd\_20XX**: Average spending per dosage unit (weighted) for the year specified.
* **Avg\_Spnd\_Per\_Clm\_20XX**: Average spending per claim for the year specified.
* **Avg\_Spnd\_Per\_Bene\_20XX**: Average spending per beneficiary for the year specified.
* **Outlier\_Flag\_20XX**: A boolean flag indicating if the data is an outlier for the year specified.
* **Chg\_Avg\_Spnd\_Per\_Dsg\_Unt\_21\_22**: Change in average spending per dosage unit from 2021 to 2022.
* **CAGR\_Avg\_Spnd\_Per\_Dsg\_Unt\_18\_22**: Compound Annual Growth Rate of average spending per dosage unit from 2018 to 2022.

**Considerations**

* Ensure the database is normalized if needed, but since all data seems related to a single entity (drug spending data per year), a single table might be sufficient.
* Data types are chosen based on typical sizes for the expected values, but they might need adjustment based on actual data ranges.

**Steps for Normalization**

1. **Identify entities and relationships**.
2. **Break the data into multiple tables** ensuring each table captures one type of entity.
3. **Establish relationships using foreign keys**.

Based on your data, we can identify the following entities:

* **Drug Information**: Basic details about the drug (brand and generic name).
* **Manufacturer**: Information about the manufacturer.
* **Yearly Data**: Annual financial and usage data for the drugs.

**Proposed Tables and Relationships**

1. **Drugs**:
   * Contains unique information about each drug.
2. **Manufacturers**:
   * Contains details about manufacturers.
3. **Drug\_Manufacturer**:
   * Manages the many-to-many relationship between drugs and manufacturers.
4. **Yearly\_Spending**:
   * Stores yearly spending and usage data related to each drug-manufacturer combination.
5. **Yearly\_Aggregate**:
   * Stores yearly aggregate statistics and derived data for a drug.

**Schema Design**

**Table: Drugs**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| Drug\_ID | INT | Primary Key, unique identifier |
| Brnd\_Name | VARCHAR(255) | Brand name of the drug |
| Gnrc\_Name | VARCHAR(255) | Generic name of the drug |

**Table: Manufacturers**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| Mftr\_ID | INT | Primary Key, unique identifier |
| Mftr\_Name | VARCHAR(255) | Name of the manufacturer |

**Table: Drug\_Manufacturer**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| Drug\_ID | INT | Foreign Key referencing Drugs |
| Mftr\_ID | INT | Foreign Key referencing Manufacturers |
| Tot\_Mftr | INT | Total number of manufacturers |
| **Primary Key**: (Drug\_ID, Mftr\_ID) |  |  |

**Table: Yearly\_Spending**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| ID | INT | Primary Key, unique identifier for each record |
| Drug\_ID | INT | Foreign Key referencing Drugs |
| Mftr\_ID | INT | Foreign Key referencing Manufacturers |
| Year | YEAR | The year of the data |
| Tot\_Spndng | DECIMAL(15,2) | Total spending for the year |
| Tot\_Dsg\_Unts | DECIMAL(15,2) | Total dosage units for the year |
| Tot\_Clms | INT | Total claims for the year |
| Tot\_Benes | INT | Total beneficiaries for the year |
| Avg\_Spnd\_Per\_Dsg\_Unt\_Wghtd | DECIMAL(15,2) | Average spending per dosage unit (weighted) for the year |
| Avg\_Spnd\_Per\_Clm | DECIMAL(15,2) | Average spending per claim for the year |
| Avg\_Spnd\_Per\_Bene | DECIMAL(15,2) | Average spending per beneficiary for the year |
| Outlier\_Flag | BOOLEAN | Indicates if the data is an outlier for the year |

**Table: Yearly\_Aggregate**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| Drug\_ID | INT | Foreign Key referencing Drugs |
| Year | YEAR | The year of the data |
| Chg\_Avg\_Spnd\_Per\_Dsg\_Unt | DECIMAL(10,4) | Change in average spending per dosage unit from previous year |
| CAGR\_Avg\_Spnd\_Per\_Dsg\_Unt | DECIMAL(10,4) | Compound Annual Growth Rate of average spending per dosage unit over the years |

**Relationships and Foreign Keys**

* **Drugs** to **Drug\_Manufacturer**: One-to-many (a drug can have multiple manufacturers).
* **Manufacturers** to **Drug\_Manufacturer**: One-to-many (a manufacturer can be associated with multiple drugs).
* **Drug\_Manufacturer** to **Yearly\_Spending**: One-to-many (a specific drug-manufacturer combination has yearly data).
* **Drugs** to **Yearly\_Aggregate**: One-to-many (each drug has yearly aggregate statistics).