Automotive Supply Chain Analysis

Lucas Lee
Department of Data Science
UCSD
San Diego, CA, 92092
l2lee@ucsd.edu

Yi Lien
Department of Data Science
UCSD
San Diego, CA, 92092
yilien@ucsd.edu

Jean Hsu
Department of Data Science
UCSD
San Diego, CA, 92092
jeh061@ucsd.edu

Abstract

Complex and interlinked supply chains define today's automotive industry. Multiple manufacturers (makers) often rely on overlapping sets of suppliers for critical vehicle components, resulting in a spiderweb of interdependencies and competitive overlaps. In this analysis, we present a detailed methodology for gathering, standardizing, and analyzing automotive supply chain data using PostgreSQL and Neo4j. In particular, we look to focus on outlining key suppliers, how they have changed over time, and how similar makers are in terms of their supply chain.

1 Introduction

1.1 Motivation

The proposed U.S. government tariffs on Canadian and Mexican imports are expected to have a significant impact on vehicle prices, potentially increasing costs by approximately $3,000 per unit [1]. Given the highly interconnected nature of the global automotive supply chain, such changes could trigger ripple effects across the industry, affecting suppliers, automakers, and consumers worldwide.

To better understand these potential disruptions, **this study aims to map out existing automotive supply chains, identifying key suppliers, their evolution over time, and the competitive landscape for automotive components and makers**. By analyzing supply chain structures and interdependencies, we can assess vulnerabilities, anticipate changes in supplier relationships, and provide insight into how the industry can respond to external pressures.

This project requires the following datasets to achieve its objectives:

- We need a graph database containing supply chain relationships, including supplier-manufacturer connections with information on which suppliers provide specific components to which automakers
- We need temporal relational data to track how these relationships have evolved over time

All the analysis in this paper can be found in the following GitHub repo.

1.2 Key questions

To investigate the interactions between suppliers, automakers, and their models within the supply network, this report focuses on three critical areas:

(1) Supply Chain Overview

- Who are the primary suppliers for each product?
- What does the key supply network look like?

(2) Time Series Analysis

- Who are the key automotive suppliers, and how have their positions evolved over time?

(3) Automaker Similarity

- How similar are automakers in their supplier relationships?
- How can this similarity be quantified?

By answering these questions, this analysis seeks to provide valuable insights into supply chain structures and trends that influence the automotive industry.

2 Data sources and preprocessing

2.1 Data Sources and data types

The primary data source for our analysis is MarkLines [2], an online database specializing in automotive industry data. This platform provides comprehensive datasets, including automotive component trading information and original equipment manufacturer (OEM) locations, etc. We extracted two types of data from the platform.

(1) Automotive component datasets

We downloaded 26 CSV files containing comprehensive supply data for various automotive components (product categories). The following are the representative automotive components we selected for our analysis:

- ABS ESC
- Automated Manual Transmission
- Automatic Transmission (AT)
- Camshaft
- Catalytic Converter
- Chassis Frame
- Clutch
- Crankshaft
- Crossmember
- Cylinder Block
- Disc Brake (Pad)
- Drum Brake (Drum)
- Exhaust Manifold
- Fuel Injection
- Fuel Pump
- Manual Transmission
- Oil Pump
- Piston
- Power Steering Pump
- Shock Absorber
- Stabilizer
- Starter Motor
- Steering Gear
- Torque Converter
- Transmission Shaft

- Water Pump

Each file contains five columns of data: the automaker's region, the automaker's name, the model name, the model year, and the supplier's name.

Each row in the dataset essentially represents a specific part (product) supplied by a supplier to a particular car model produced by a manufacturer (maker).

(2) Maker's group

Additionally, we extracted manufacturer group data from MarkLines to facilitate future mapping. This dataset includes the names of automakers and their corresponding manufacturer groups.

2.2 Data preprocessing

Our raw data has been preprocessed through the following steps:

(1) Combine datasets and assign trading volumes

Our dataset originated from separate CSV documents, each capturing (Region, Maker, Model, Supplier) tuples for a specific vehicle component. The first step in preprocessing is to create a master table that includes all the tuples across all products following the steps:

1. We concatenated all product-specific CSV files row-wise to create a unified master table, preserving all original relationships between makers, models, and suppliers
2. We add a 'product' column to each row, indicating which automotive component (e.g., transmission, brake system, electrical system) the supplier-maker relationship represents. This column was populated based on the source CSV file name for each row
3. We add a 'volume' column that models the yearly number of units supplied in each maker-model-supplier relationship. These values were generated by sampling from $U(100, 1000)$ to simulate varying production scales across different vehicle models.

This was done through Pandas and Python, with code available at A.1

(2) Standardize model region data

In our dataset, model names include associated region information; however, the format of this information is inconsistent. Some regions are represented using ISO country codes (e.g., 'USA'), while others are written in their full names (e.g., 'Mexico'). To standardize this, we referenced a country mapping table to convert all region names to their full official format.

After extracting and standardizing the region data, we appended it as 'model_region' to the dataset.

(3) Extract supply group data from supplier data

We observed that many supplier names share common prefixes, often indicating affiliation with a parent company. For example, 'AAPICO Amata Co., Ltd.' and 'AAPICO Hitech Public Co., Ltd.' both belong to AAPICO. To ensure a clearer analysis and more accurate market representation, we created a supplier table to store parent company information for further analysis.
To systematically identify these relationships, we:

- **Identify Common Prefixes**
  - Using regexp_match(name,`\w+\s`) to extract the first word from each supplier name
  - Counting occurrences of each extracted first word across all suppliers
  - Selecting only those prefixes that appear in more than one supplier name (HAVING COUNT(*) > 1)
- **Find Matching Supplier Names**
  - Cross joining suppliers with the parent CTE to compare all names against the extracted prefixes
  - Using WHERE LOWER(name) `~*`LOWER(parent.match) to find supplier names that contain the identified prefix
  - Ensuring that the name actually starts with the prefix using "LEFT"(LOWER(name), LENGTH(match)) = LOWER(match)

  - Trimming any trailing spaces from the prefix (LEFT(match, LENGTH(match)-1) AS match_name)

- **Assign The Matched Prefix as parent_company in The Supplier Table**
  - Updating the supplier.parent_company column
  - Joining with the match CTE to ensure each supplier receives the correct parent company mapping

```sql
with parent as (
select distinct match,count(*) from (
    select supplier_id,name,unnest(regexp_match(name,'\w+\s')) as match
    from supplier
              ) s
group by match
having count(*)>1),
match as (
select name, "left"(match,length(match)-1) as match_name from supplier
cross join parent
where lower(name) ~* lower(parent.match) and "left"(lower(name),length(match)) = lower(match)
order by lower(match))

update supplier
set parent_company = match_name
from match
where supplier.name = match.name;
```

Figure 1: Query: Create Mapping to Supply Group

- **Manually Verify Parent Companies**
  Some prefixes, such as city names like Shanghai or Beijing, were incorrectly grouped as parent companies due to their frequent appearance in Chinese supplier names. To address this, we manually verified parent companies using official websites and refined the mappings accordingly.

```sql
update supplier
set parent_company ='ADAC'
where name like 'ADAC Automotive%'
update supplier
set parent_company = 'AGS'
where name like 'AGS %'
update supplier
set parent_company = 'ATA Casting Technology'
where parent_company = 'ATA'
update supplier
set parent_company = null
where parent_company in ('AT','Changchun','Changsha','Chengdu','China','Chongqing','Foshan','Fuxin','GF','Guangdong','Hangzhou'
,'Harbin','Hebei','Hiroshima','Hubei','Hunan','Jiangsu','Jinzhou','Liuzhou','Lucas','Maxion','Metal','Thai','The','Tianjin'
,'Nanjing','New','Ningbo','Precision','SKH','Shandong','Shanghai','Shenyang','Siam','Sichuan','Taiwan','Wenzhou','Wuhan'
,'Wuhu','Wuxi','Yantai','Yuhuan','Zhejiang')
or name in ('American Axle & Manufacturing Holdings, Inc.','Art Metal Mfg. Co., LTD.','Asahi Tec Aluminium (Thailand) Co., Ltd.',
           'Asahi Tekko Co., Ltd.','Asama Giken Co., Ltd.')
or supplier_id in ('125','126','130','134','142','143','309','310','311')
update supplier
set parent_company = 'Daihatsu'
where name = 'Akashi-Kikai Industry Co., Ltd.'
update supplier
set parent_company = 'Akebono Brake Industry'
where name like 'Akebono%'
update supplier
set parent_company = 'Allison Transmission'
where parent_company ='Allison'
update supplier
set parent_company = 'American Mitsuba Corp.(AMC)'
where name like 'American Mitsuba%'
```

Figure 2: Query: Manually Update Supply Group (partial)

3 Methodology and Demonstration

3.1 Supply chain overview

(1) Trading volume analysis

In this section, we want to find the top five suppliers in each product category in a given time frame to better understand the key suppliers over time.

We began by joining the supplier group data into the preprocessed table using PostgreSQL. Since the synthesized volume fluctuates yearly, we categorized the years into five-year intervals to aggregate trading volumes. Finally, we generated a table displaying the top five suppliers for each product in the given year range. The output table serves as a reference to help us gain a broad understanding of the key suppliers. An example of the output is as follows:

product	year_range	supplier_group	total_volume
abs_esc	2005-2010	ADVICS	920
abs_esc	2011-2015	Bosch	48934
abs_esc	2011-2015	Continental	23024
abs_esc	2011-2015	ZF	12420
abs_esc	2011-2015	Beijing West Industries	8838
abs_esc	2011-2015	ADVICS	4660
abs_esc	2016-2020	Bosch	573248
abs_esc	2016-2020	Continental	313680
abs_esc	2016-2020	ZF	97246

Figure 3: Example output of the trading volume analysis

(2) Key supply network analysis

In this analysis, we aim to find the key supply network by extracting the leading suppliers in each product category over the years.

We first load the preprocessed data in PostgreSQL and follow a similar approach to the previous analysis. Then, we identify the top suppliers in each product category and visualize the corresponding supply network using Neo4j. In the following analysis, we will illustrate how we select the top three suppliers in the 'transmission shaft' category.

First, we load the PostgreSQL table into Neo4j. We then merge nodes and edges, setting the supplied product as an attribute on the edges to enable easier extraction later. Next, we focus on the transmission shaft category, aggregating either trade count or total trading volume to identify the top three suppliers. Then, we match the edges related to them within a specified year range for transmission shafts. Finally, we create virtual edges for visualizing the supply network, returning the makers, the top three suppliers, and their connections as the final output. The following are the details for the output of the analysis.

In Figure 4, the explanation of elements is as follows:

- Green nodes represent automakers, which corresponds to the 'maker' in our data.
- Brown nodes represent suppliers, which corresponds to the 'supplier_group' in our data.
- Edges indicate that there is a "SUPPLIES_TO" relationship, where a supplier provides a specific automotive component to an automaker in the given year range.
- Edge colors differentiate supply relationships across different time frames, with red (2005-2010), yellow (2011-2015), green (2016-2020), and blue (2021-2025).

In Figure 5, the elements are similar to the previous explanation, with only the thickness of the edges adjusted to reflect the trade count.

In Figure 6, the elements are similar to the element explanation in Figure 4. However, noted that the top 3 suppliers are selected based on total trading volume instead of trade count.

All the network visualizations highlight the significant transformation in the automotive supplier ecosystem from 2005 to the landscape of 2025.

- **Early Period (2005–2010)**: Neumay dominated the transmission shaft market, primarily supplying European manufacturers.
- **Transition Period (2011–2015)**: Bosch expanded its presence by establishing relationships with Chinese automakers such as SAIC, Dongfeng, and BAIC, while Qianchao Sunway

Co., Ltd. developed connections with joint ventures like SAIC Volkswagen and FAW-Volkswagen.

- **Mature Period (2016–2020)**: SKF emerged as a key hub, forming extensive connections with manufacturers in the United States and Japan.

- **Recent Period (2021–2025)**: Japanese supplier Aichi has strengthened its relationships with domestic manufacturers, including Toyota and Mitsubishi, indicating a shift toward more specialized Asian supplier networks.

This analysis reveals a clear geographic shift in the transmission shaft supply chain over time. The transition from Neumay to SKF reflects manufacturers' evolving sourcing strategies and efforts to diversify trading relationships, particularly for North American brands. Initially, European suppliers primarily served European manufacturers. Over time, Chinese joint ventures integrated into the supply chain, and in recent years, Japanese suppliers have entered the market, competing with dominant players over Asian automakers.

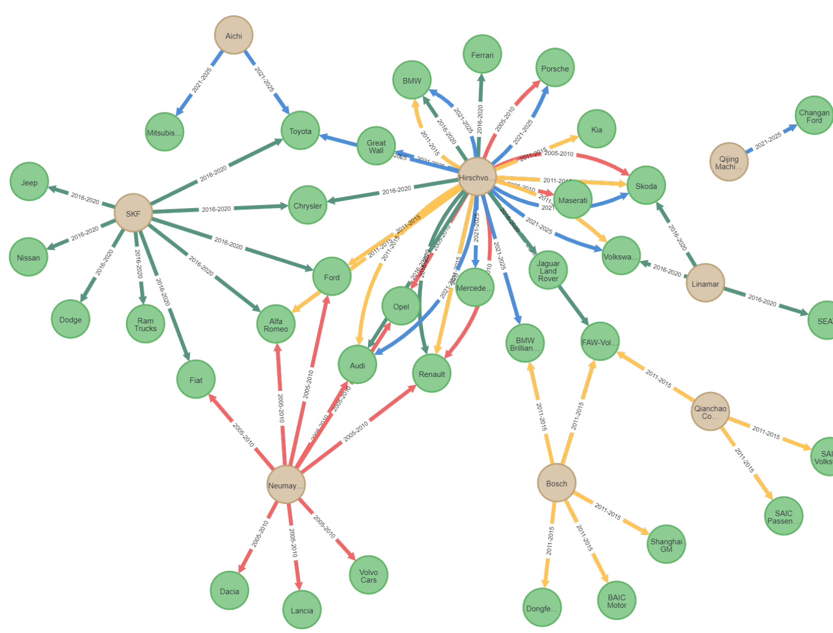

Figure 4: Top 3 Suppliers based on Trade Year Count in Transmission Shaft over time, with edge color representing time

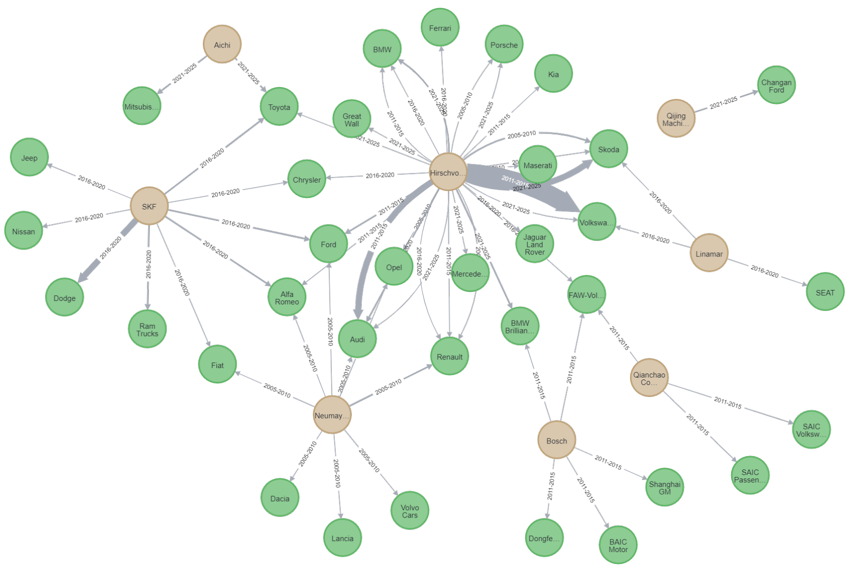

Figure 5: Top 3 Suppliers based on Trade Year Count in Transmission Shaft over time, with edge thickness representing count

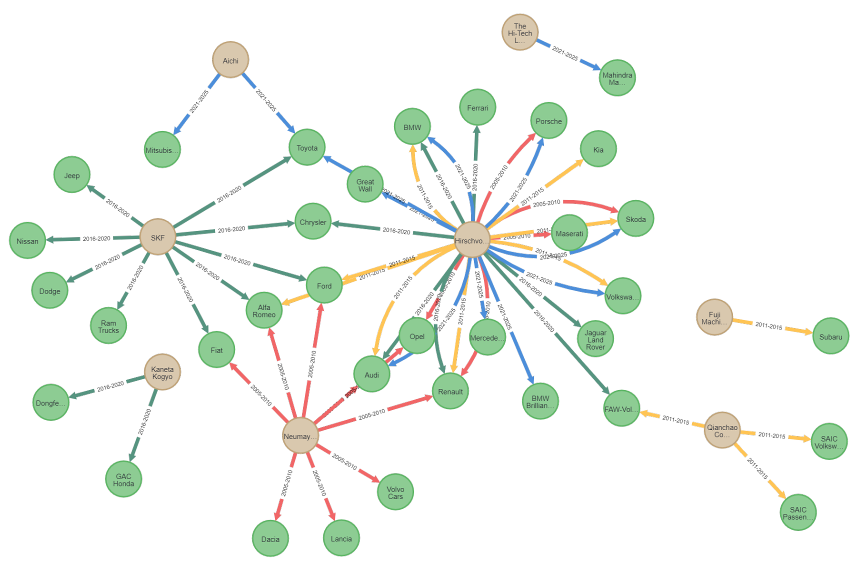

Figure 6: Top 3 Suppliers based on Total Trading Volume in Transmission Shaft over time

3.2 Time series analysis

To analyze market trends and uncover relationships between different suppliers and components over time, we utilize PostgreSQL for time series analysis and the definition of new metrics.

(1) 2025 Top 10 Automotive Suppliers

- Identify 2025 Top 10 Suppliers

  Our primary goal is to examine the trend among the top 10 automotive suppliers in 2025. To achieve this, we first structure a query to identify 2025 top 10 suppliers:

- **Rank Suppliers by Sales Volume**
  We assign each supplier a rank based on their annual sales volume across different products. This ranking is computed using the RANK() window function and stored in a Common Table Expression (CTE) called s_rank.
- **Calculate Year-over-Year (YoY) Trends**
  To assess changes over time, we create another subset using a LEFT JOIN, linking each supplier's data from one year to the previous year. This allows us to compute the YoY percentage change in sales for each supplier.
- **Filter for the Top 10 Suppliers in 2025**
  Finally, we filter the results to include only suppliers ranked in the top 10 for 2025. This enables us to focus on the dominant players and analyze their market trends.

This method allows us to identify key suppliers with sustained market dominance, track their performance trends, and analyze shifts in market share effectively.

```sql
with s_rank as (
select
    model_year,
    coalesce(s.parent_company,s.name) as supplier_name,
    rank() over(partition by model_year order by sum(unit) desc) as supplier_rank,
    sum(unit) as unit
from component_fact c
left join supplier s on c.supplier = s.name
group by model_year,supplier_name),
s_share as (
select
    r1.model_year,
    r1.supplier_name,
    r1.supplier_rank,
    r1.unit,
    sum(r1.unit) over(partition by r1.model_year) as unit_year,
    concat(round(cast(r1.unit / sum(r1.unit) over(partition by r1.model_year)*100 as numeric),2),'%') as share,
    case
        when r1.model_year = 2025 then concat(round(cast((r1.unit*12/3 - r2.unit) / r2.unit as numeric)*100,2),'%')
    else concat(round(cast((r1.unit - r2.unit) / r2.unit as numeric)*100,2),'%')
    end as yoy
from s_rank r1
left join s_rank r2 on r1.model_year = r2.model_year+1 and r1.supplier_name = r2.supplier_name)
select model_year, supplier_name, supplier_rank,share ,yoy
from s_share
where
    model_year =2025 and supplier_rank <= 10
order by model_year,supplier_rank;
```

Figure 7: Query: Identify 2025 Top 10 Suppliers

From the result of the query, we observed:

- Bosch leads the market, contributing nearly a quarter of the total supply chain.
- Faurecia and FinDreams demonstrate significant YoY growth, nearly tripling their sales from the previous year.

model_year	supplier_name	supplier_rank	share	yoy
2025	Bosch	1	25.64%	-18.94%
2025	Continental	2	7.57%	-43.63%
2025	Hitachi	3	6.74%	21.99%
2025	Faurecia	4	5.28%	307.72%
2025	FinDreams	5	5.22%	307.11%
2025	Wuhu Bethel Electronic Control System Co.,…	6	4.44%	-22.03%
2025	ZF	7	3.08%	-65.52%
2025	Tenneco	8	2.48%	74.49%
2025	Vitesco	9	2.27%	297.88%
2025	Global Technology Co., Ltd.	10	2.25%	241.94%

Figure 8: Result: Identify 2025 Top 10 Suppliers

- Drivers Behind Faurecia's and FinDreams's High YoY Growth

  A closer examination of Faurecia and FinDreams reveals that their impressive YoY growth in 2025 is largely driven by strong collaborations with Chinese and Indian manufacturers. Please see the code at B.1

	parent_company	region	model_region	product
1	Faurecia	ASEAN, India, Korea	India	exhaust_manifold
2	FinDreams	China	China	abs_esc

Figure 9: Result: Supplier Collaborations by Region and Product

- Track Market Trends of 2025's Top 10 Suppliers Over Time

  To analyze the evolution of the top 10 suppliers in 2025, we refine our query by adjusting the WHERE clause to retrieve their yearly data across all available years. For detailed query, please check B.2

  Given the large dataset, we visualize the results in Excel to better interpret market trends and patterns.

Output | Time Series Analysis 2025 Top 10

	model_year	supplier_name	supplier_rank	share	yoy
1	2008	Bosch	3	6.79%	<null>
2	2008	ZF	6	3.66%	<null>
3	2008	Hitachi	17	1.47%	<null>
4	2008	Continental	36	0.52%	<null>
5	2008	Tenneco	56	0.14%	<null>
6	2009	ZF	5	4.98%	10.75%
7	2009	Bosch	9	4.19%	-49.85%
8	2009	Hitachi	31	0.65%	-63.99%
9	2009	Continental	49	0.19%	-69.34%
10	2010	Hitachi	5	6.34%	2097.52%
11	2010	Bosch	8	3.66%	96.87%
12	2010	ZF	9	2.49%	12.55%
13	2010	Continental	27	0.45%	418.62%
14	2011	Bosch	2	6.05%	898.08%
15	2011	Hitachi	3	5.04%	380.39%

119 rows

Figure 10: Result: Market Share Trends of 2025's Top 10 Suppliers Over Time (partial)

- Market Share Trends
  The chart reveals key trends in supplier market share over time.

  - **Bosch (orange)**
    The orange bars represent Bosch, which saw a decline around 2012 but rebounded after 2015, steadily increasing its share to 26% in 2025, making it a dominant player in the supply chain.

  - **ZF (purple)**
    ZF has consistently maintained a significant presence. In 2012, as Bosch's share declined, ZF captured a larger portion of the market. However, beyond that peak, its share remained stable at around 5% per year without major fluctuations.

  - **Continental (dark green) & Global Technology (light blue)**
    Continental and Global Technology began to gain traction after 2015, showing a noticeable upward trend.

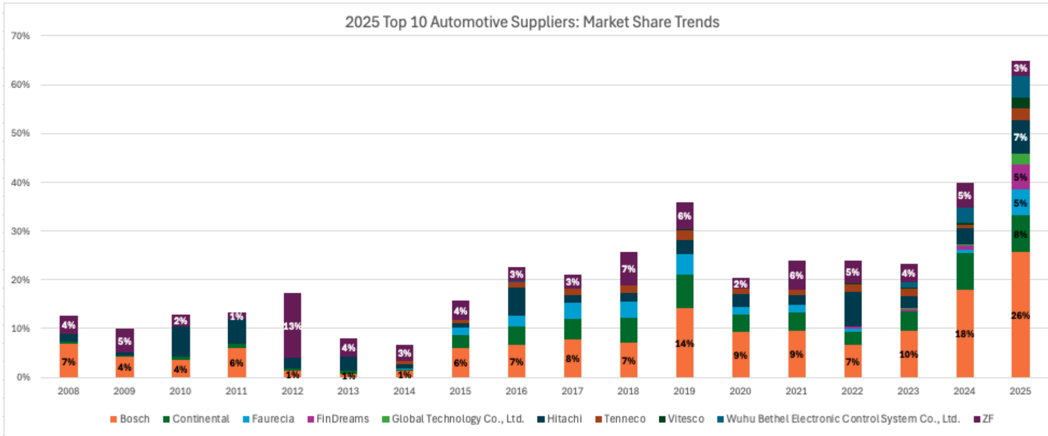

Figure 11: Visualization: Market Share Trends

(2) Supplier-Automaker Collaborations

- Average Maker Amount
  To explore which companies collaborate with the most different automakers, we designed a query to analyze the number of distinct automakers each supplier works with annually.
  - **Count Distinct Automakers per Year**
    For each supplier and model year, we count the number of unique automakers they supply to using COUNT(DISTINCT maker).
  - **Calculate the Average Maker Count per Supplier**
    We compute the average yearly number of automakers per supplier by dividing the total automaker collaborations (SUM(maker_num)) by the number of distinct years the supplier appears (COUNT(DISTINCT model_year)).
  - **Group by Parent Company**
    To ensure accurate supplier aggregation, we use COALESCE(parent_company, name), which groups suppliers under their respective parent companies where applicable.
  - **Rank the Top 10 Suppliers**
    Finally, we order the results by average maker amount in descending order and limit the output to the top 10 suppliers with the widest automaker partnerships.

```sql
select
    coalesce(parent_company,name) as supplier,
    sum(maker_num) / count(distinct model_year) as avg_maker_amount
    from (
select distinct model_year, c.supplier, count(distinct maker) as maker_num from component_fact c
group by model_year, supplier
order by model_year, supplier) c
left join supplier s on c.supplier = s.name
group by coalesce(parent_company,name)
order by avg_maker_amount desc
limit 10
```

Figure 12: Query: Average Maker Amount

- Findings and Industry Impact

  - Bosch and ZF have the highest number of automaker partnerships, which aligns with their strong market position in the automotive supply chain. Their extensive collaborations may contribute to their resilience and influence in the industry.
  - Continental, Aisin, and Faurecia also demonstrate a broad range of partnerships with different automakers, reinforcing their competitive presence.

These findings closely align with the market share trends observed earlier, suggesting that suppliers with a wider range of automaker collaborations tend to hold a stronger position in the industry. Understanding these relationships helps identify key players with extensive influence across multiple manufacturers.

	supplier	avg_maker_amount
1	Bosch	47.666666666666667
2	ZF	32.7222222222222222
3	Continental	26.166666666666667
4	Aisin	22.4705882352941176
5	Faurecia	18.3333333333333333
6	Precision Sintered Products (Wuxi) Co., Ltd.	16.5
7	Zhejiang Qibo Machinery Co., Ltd.	15
8	Denso	14.2222222222222222
9	Chengdu Xiling Power Science & Technology Incorpor…	14
10	Hitachi	13.7222222222222222

Figure 13: Result: Average Maker Amount

(3) Case Study: Bosch & ZF

Following our analysis of the top 10 automotive suppliers, Bosch and ZF emerged as two of the strongest performers over time. To gain deeper insights into their market strategies, we further analyzed their regional distribution and product focus trends.

This analysis builds upon our previous findings by refining the dataset to focus exclusively on Bosch and ZF. We adjusted the WHERE clause to filter for these two suppliers and applied a CASE statement to align model regions with broader maker region segments for consistency in regional analysis.

This query aims to analyze Bosch and ZF's supplier trends, focusing on three key aspects:

- **Maker Regional Distribution:**
  Observes how Bosch and ZF's automaker partnerships have evolved across different regions.
- **Model Regional Distribution:**
  Shows where their supplied components are used in vehicle production and which regions the vehicles are intended for.
- **Product Focus:**
  Highlights how Bosch and ZF have shifted their manufacturing strategies across different automotive components.

This query B.3 provides a comprehensive view of Bosch and ZF's supply chain strategies, allowing for deeper insights into their regional partnerships, component distribution, and product specialization over time.

- (3.1) Bosch
  - **Changes in Maker Regional Distribution**

    The chart here illustrates the regional distribution of automakers that Bosch collaborates with over time.
    * The loss of cooperation with American automakers in 2012:
      The blue line represents Bosch's overall market share, and we notice the significant drop in 2012 might be related to the loss of partnerships with American automakers, as we can see the blue bar representing Americas disappeared in 2012.
    * Shift in Focus from Europe to China in 2015 (onward)
      By looking at the light blue part, we can see Bosch initially had a strong focus on European automakers; however, starting in 2012, Bosch gradually shifted its focus toward Chinese automakers, as reflected in the increasing proportion of the China segment in the bars.

This strategic shift appears to be effective, as Bosch's market share has consistently grown alongside its increasing collaborations with Chinese automakers. This analysis suggests that Bosch's decision to expand partnerships in China has played a key role in its market growth over time.

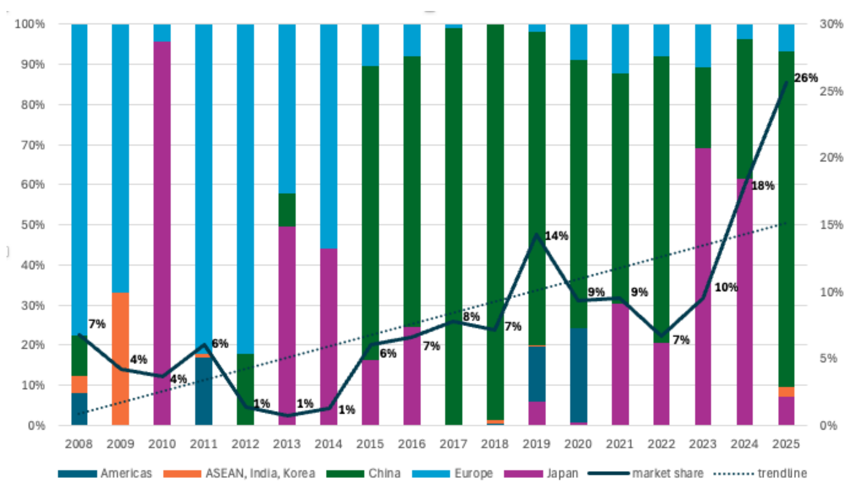

Figure 14: Visualization: Bosch's Changes in Maker Regional Distribution

- **Changes in Model Regional Distribution**

  * Similar trends with maker distribution: Shift their business focus from Europe to China
    Bosch's model regional distribution follows a similar trend to its maker distribution, reflecting a strategic shift. Initially, Bosch's components were largely tied to European vehicle production, but from 2012 onward, the China segment grew significantly.

This shift aligns with Bosch's increasing partnerships with Chinese automakers, suggesting a response to rising demand in the Chinese market. While the chart does not directly show component usage, the parallel trends indicate that Bosch's regional focus adjustment contributed to its market expansion.

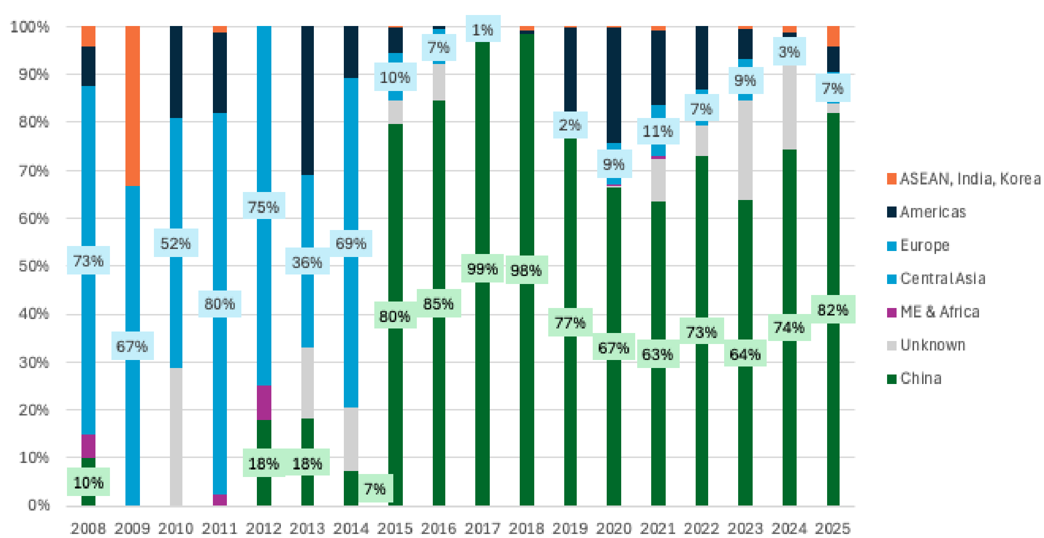

Figure 15: Visualization: Bosch's Changes in Model Regional Distribution

– **Changes in Product Focus**

∗ Transition from Fuel Injection and Starter Motor to ABS/ESC
The chart shows that Bosch gradually shifted its focus from fuel injection and starter motors to ABS and ESC starting in 2015.

We might found that between 2012 and 2015, Bosch not only adjusted its automaker partnerships, but also transformed its product strategy. This dual shift in both market focus and product offerings likely played a crucial role in its continued growth and market dominance.

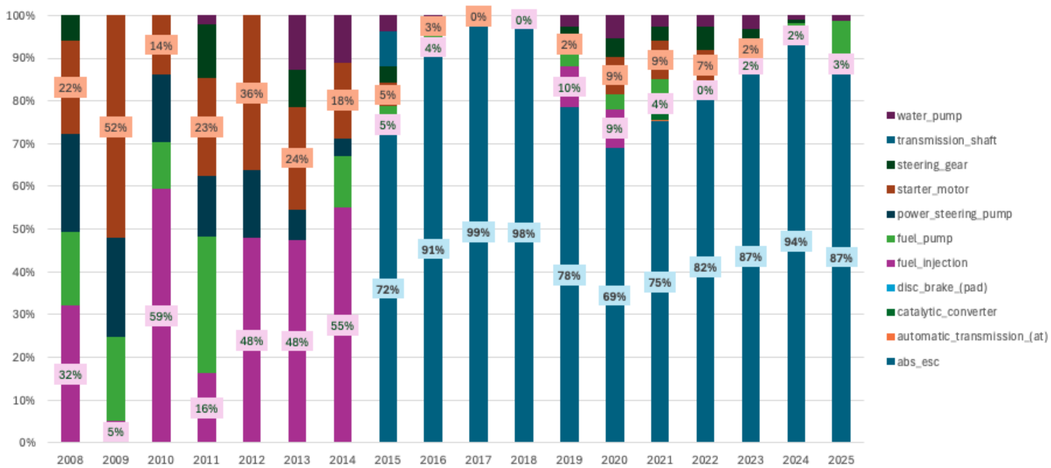

Figure 16: Visualization: Bosch's Changes in Product Focus

- (3.2) ZF

– **Changes in Maker Regional Distribution**

∗ Maintain High Participation in Europe
ZF has consistently maintained a strong presence in Europe, indicating its long-standing partnerships with European automakers.

∗ Growth in China and Japan
Despite the stability in Europe, ZF has also expanded its collaborations with Chinese and Japanese automakers over time.

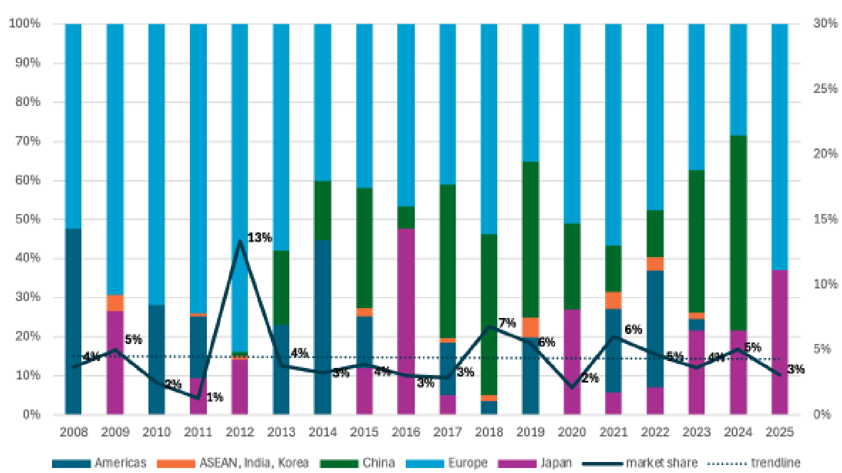

Figure 17: Visualization: ZF's Changes in Maker Regional Distribution

– **Changes in Model Regional Distribution**

  * Maintain a Strong Focus on Europe and The Americas
    The chart shows that European vehicle production consistently represents a significant portion of ZF's model distribution, reinforcing its deep-rooted partnerships with European automakers.

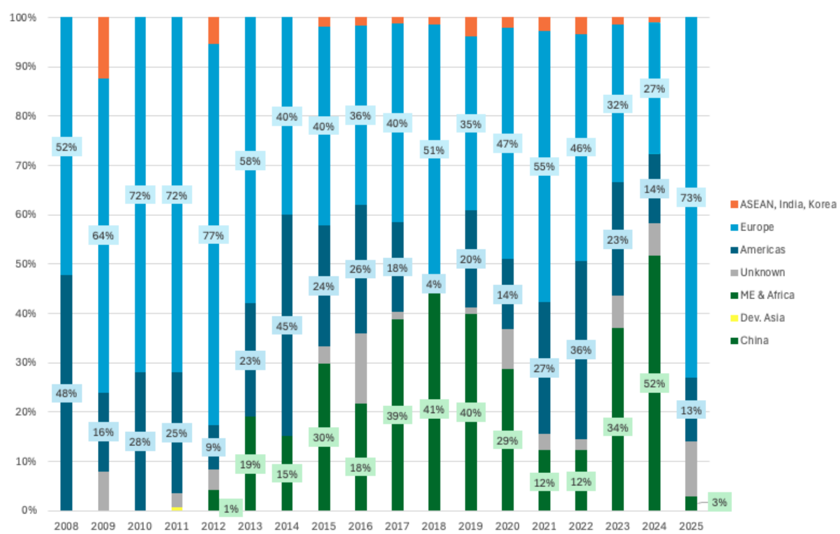

Figure 18: Visualization: ZF's Changes in Model Regional Distribution

– **Changes in Product Focus**

  * Transition from Clutch and Steering Gear to Shock Absorber and Transmission
    we can see a transition from clutch and steering gear to shock absorbers and transmissions over time. Additionally, ZF's product portfolio appears more diverse compared to Bosch.
  * Collaborative relationship with Bosch
    Another key observation is that Bosch and ZF do not have overlapping product selections, suggesting that rather than being direct competitors among those components, they may actually be in a collaborative relationship within the automotive supply chain.

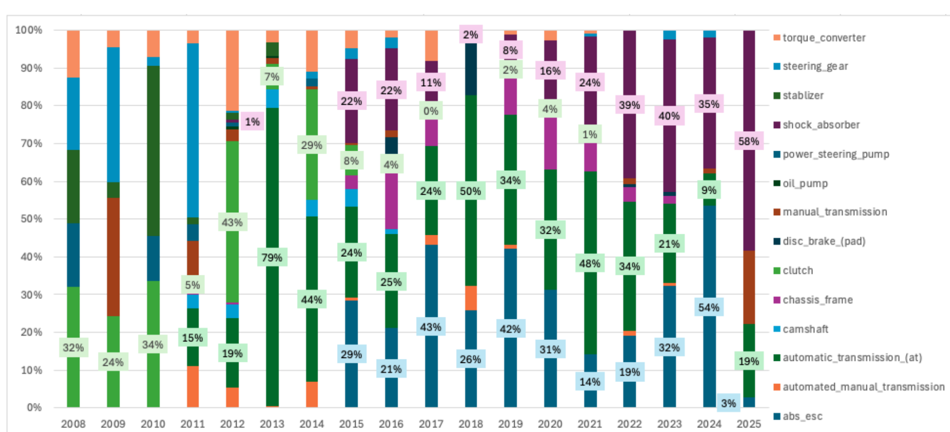

Figure 19: ZF's Changes in Product Focus

(4) Top 3 suppliers in each product category
In this section, we further analyze the dominant suppliers in each product category by identifying those that consistently rank among the top 3 over multiple years.
Initially, we attempted to track market share changes for each component, but supplier rankings fluctuated significantly year over year, making it difficult to establish clear trends. To address this, we counted how many times each supplier appeared in the top 3 for a given product each year and ranked them based on their total appearances. This method provides a more stable measure of long-term dominance rather than focusing on short-term variations. This approach helps us identify suppliers with a strong and sustained presence in the market. Additionally, we created an array that lists the specific years and products where a supplier held a top 3 position. The detailed query could be seen in B.4

Our analysis highlights suppliers with sustained market leadership. For example, Aisin ranks highest in total top 3 appearances, reinforcing its strong and consistent supply across multiple product categories. In 2008, Aisin secured top 3 positions in manual transmission, automatic transmission, and oil pump, demonstrating its broad industry presence.

	supplier_name	high_rank_count	high_rank_products
1	Aisin	65	{2008:manual_transmission,2008:automatic_transmi…
2	ZF	61	{2008:steering_gear,2008:stablizer,2008:power_st…
3	Bosch	55	{2008:power_steering_pump,2008:fuel_injection,20…
4	Denso	38	{2008:starter_motor,2009:fuel_pump,2010:fuel_pum…
5	Hitachi	38	{2008:piston,2008:shock_absorber,2009:shock_abso…
6	Daihatsu	24	{2011:manual_transmission,2011:drum_brake_(drum)…
7	MAHLE	24	{2008:piston,2008:camshaft,2009:piston,2010:pist…
8	JTEKT	23	{2009:steering_gear,2010:power_steering_pump,201…
9	Hyundai	20	{2008:water_pump,2014:manual_transmission,2014:a…
10	KYB	20	{2008:shock_absorber,2009:shock_absorber,2009:po…

Figure 20: Result: Top 3 Suppliers in Each Product Category (partial)

(5) Component Competition Analysis
In this section, we analyze the competitiveness of different automotive components by examining the number of suppliers for each product category over time.

- **Methodology**
  To measure competition, we count the number of distinct suppliers per year for each product, then calculate the average number of suppliers per product and rank them accordingly. This approach allows us to identify components with high supplier diversity versus those dominated by a few key players.

```
with competition as (
select distinct model_year, product,
                count(supplier) over(partition by model_year, product) as company_total
from (select model_year, product, supplier from component_fact
        group by model_year, product, supplier) c
group by model_year, product,supplier
order by model_year)
select distinct product, avg, rank()over(order by avg desc) from
(select distinct product, avg(company_total) over(partition by product) as avg from competition) as rank
order by rank
```

Figure 21: Query: Component Competition Analysis

- **Demonstration**

  - The catalytic converter has the highest number of suppliers, suggesting a highly competitive market or that this component requires fewer specialized technologies, making it easier for multiple suppliers to enter.
  - On the other hand, the transmission shaft has the lowest number of suppliers, with an average of only four. This indicates a market with high barriers to entry, potentially making it an oligopoly where only a few dominant suppliers operate.

	product	avg	rank
1	catalytic_converter	65.7272727272727273	1
2	abs_esc	36.2307692307692308	2
3	chassis_frame	20.6111111111111111	3
4	shock_absorber	20.0952380952380952	4
5	automatic_transmission_(at)	18.3333333333333333	5
6	crossmember	15.1875	6
7	water_pump	14.6111111111111111	7
8	clutch	13.9411764705882353	8
9	starter_motor	13.8333333333333333	9
10	camshaft	12.8333333333333333	10
11	manual_transmission	12.8333333333333333	10
12	crankshaft	12.2941176470588235	12
13	fuel_pump	12	13
14	disc_brake_(pad)	11.1764705882352941	14
15	exhaust_manifold	11.1666666666666667	15
16	steering_gear	10.7222222222222222	16
17	oil_pump	10.5555555555555556	17
18	stablizer	10.2352941176470588	18
19	fuel_injection	10.1666666666666667	19
20	cylinder_block	9.9411764705882353	20
21	piston	8.7222222222222222	21
22	torque_converter	7.3125	22
23	power_steering_pump	6.3125	23
24	automated_manual_transmission	5.1764705882352941	24
25	drum_brake_(drum)	5	25
26	transmission_shaft	4.1111111111111111	26

Figure 22: Result: Component Competition Analysis

3.3 Maker's similarity

Much of the previous sections have focused on the supply chain in the automotive industry in terms of suppliers. Here, we shift perspectives to view the supply chain in terms of makers and consumers.

The main goal of this section is to answer the following questions:

- How similar are the supply chains of different automakers?
- How can we quantify this similarity based on our data model?

3.3.1 Data Model

To turn our relational data in 2.2, to a graphical one in Neo4i, we use the following piece of code in C.1.

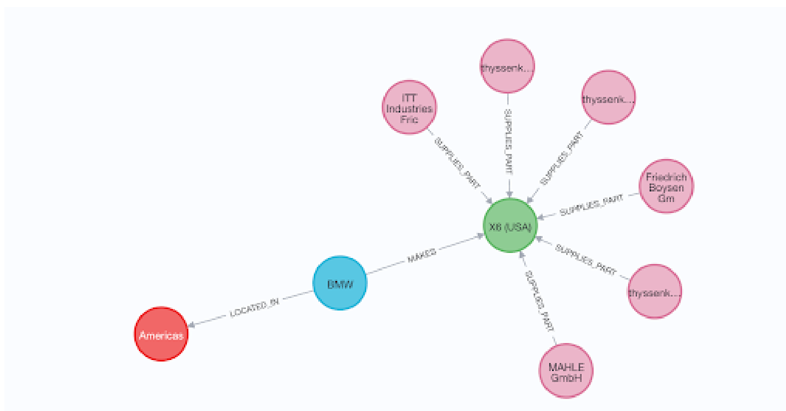

Figure 23: Data model used in automotive maker analysis.

The result of the data model can be seen in Figure 23. In this model, there are multiple types of nodes and types of edges.

Node types are separated into

- Region
- Maker
- Model
- Supplier

Further, it takes on the following relationships with one another: (Region) <-[LOCATED_IN]- (Maker) -[MAKES]-> (Model) -[USES]-> (Product) <-[SUPPLIES]- (Supplier).

What this essentially means is that

- Regions contain makers
- Makers produce models
- Models use products
- Suppliers supply products

A more complicated view of the data model can be shown in Figure 24.

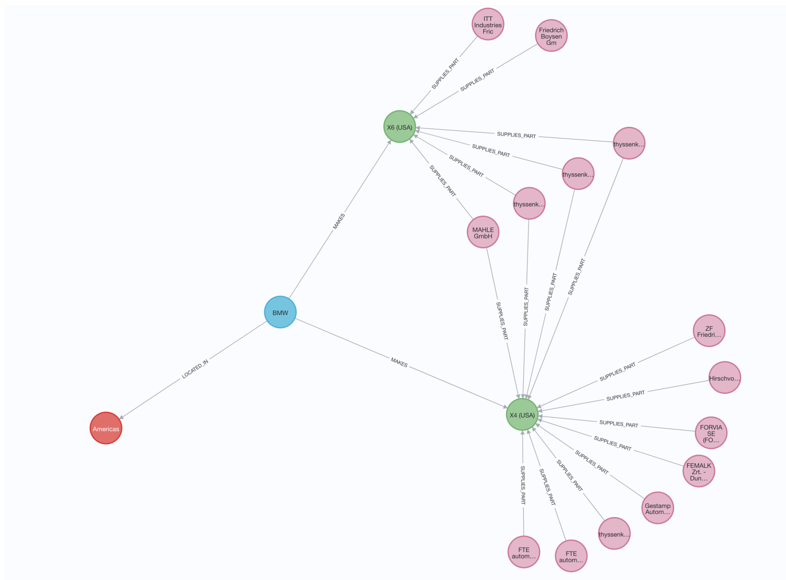

Figure 24: Maker's data model including two models from the same maker, with several shared makers

3.3.2 Similarity Metrics

As mentioned in our data model, each **maker** can have multiple **models**, and each model may be associated with one or more **parts** that come from various **suppliers**. Because of this many-to-many structure, it is useful to think of each maker's entire "supplier network" as a single set of suppliers—across all its models and parts—and then compare these sets between different makers.

Global Jaccard Similarity
To begin, we define the *global* supplier set for Maker i, S_i, as all suppliers providing any part to any model for Maker i

This means that similarity between two makers can be done by comparing their respective supplier network across all its models and parts. This can be done by computing the Jaccard similarity between two makers A and B as follows:

$$\text{Jaccard}(A, B) = \frac{|S_A \cap S_B|}{|S_A \cup S_B|} \tag{1}$$

A higher Jaccard score (closer to 1) indicates that the two makers rely on very similar sets of suppliers, often signaling strong competition or overlapping supply-chain strategies. Conversely, a lower score (near 0) suggests little overlap in their supply bases.

In practice, we implement this calculation in Neo4j by retrieving the supplier sets for each maker and then performing pairwise intersection and union operations. (See Section C.2 for the code reference)

Product-Level Jaccard Similarity
While the global Jaccard similarity is a powerful measure of overall supplier overlap, some questions require a *component-specific* view. For example, it may be more insightful to know how similar two automakers are specifically in their engine suppliers, rather than across all parts.

To achieve this, we *filter or group* by a particular product category before computing Jaccard. That is, for each product type p (e.g., "engine," "transmission," "brake_system"), we define $S_{i,p}$ as all suppliers that provide product p to Maker i.

This jaccard similarity analysis can be extended to

$$\text{Jaccard}(A, B, p) = \frac{|S_{A,p} \cap S_{B,p}|}{|S_{A,p} \cup S_{B,p}|}, \tag{2}$$

Thereby focusing on how similar Makers A and B are for *product* p. This gives us a fine-grained view of whether two manufacturers share suppliers within each product category (e.g., do they use the same brake-system suppliers?).

This is also implemented in Neo4j, see Section C.3 for the code reference).

Weighted Jaccard Similarity
Finally, not all components are equally critical or costly. An engine, for example, might be far more significant (in terms of cost or safety) than an interior trim piece. To account for this, we *weight* each product category by its relative importance:

$$\begin{aligned} J_{\text{weighted}}(A, B) &= \sum_p w_p \times \frac{|S_{A,p} \cap S_{B,p}|}{|S_{A,p} \cup S_{B,p}|} \\ &= \sum_p w_p \times \text{Jaccard(A, B, p)} \end{aligned} \tag{3}$$

where w_p is a weight (e.g., based on cost, safety criticality, or strategic value) for product p. This approach lets us highlight overlaps in *key* components—such as powertrains or braking systems—while de-emphasizing less significant overlaps (e.g., minor interior features).

Systematically, this is just done through feeding the outputs CSV of the output of Neo4j query to a Python/Jupyter notebook calculate the weighted difference. (See Section C.4 for the code reference).

In practice, domain experts and users of our code can tailor the weights w_p to reflect specific business goals or concerns (e.g., focusing on safety-critical components for risk analysis). Thus, the progression from a *global* Jaccard measure to a *product-level* and then *weighted* Jaccard measure provides a flexible, multi-scale way of analyzing how—and in which parts—two automakers' supply chains intersect.

3.3.3 Demonstration

Global Jaccard Analysis
When we run the global jaccard similarity score for every pair of makers across all regions, a few pattern emerge. From Figure 25, we see that the top 8 pairs have fairly high Jaccard similarity of over

Maker1	Maker2	jaccardSimilarity	intersection
FAW Toyota	GAC Toyota	0.5689655172413790	33
Dongfeng Honda	GAC Honda	0.5625	27
Citroen	Peugeot	0.5595238095238100	47
SAIC GM	Shanghai GM	0.53125	34
Hyundai	Kia	0.45794392523364500	49
Beijing Hyundai	Dongfeng Yueda Kia	0.425531914893617	20
Opel	Peugeot	0.41304347826087000	38
Porsche	Skoda	0.410958904109589	30
Mahindra & Mahindra	Tata	0.39285714285714300	22
Bentley	Lamborghini	0.38461538461538500	10
Audi	Porsche	0.3805309734513270	43
Fiat	Peugeot	0.375	30
Audi	Volkswagen	0.37423312883435600	61

Figure 25: Top 13 most similar automakers by global Jaccard similarity

0.4 with a non-trivial amount of shared makers.

In reality, these are all either joint ventures or those who are under a common parent company. For instance, FAW Toyota & GAC Toyota or Dongfeng Honda & GAC Honda frequently exceed Jaccard scores of 0.56, reflecting extensive shared supplier networks. Meanwhile, automakers connected by a common parent company, such as Citroen & Peugeot under the PSA Group or Hyundai & Kia under the Hyundai Motor Group, often display moderately high overlaps as well. This is consistent with the notion that corporate linkages typically centralize R&D and procurement, thus increasing supplier-base convergence.

Thus, from the use of Jaccard similarity we can see that it highlights makers who are in close relation to one another.

Maker1	Maker2	jaccardSimilarity	intersection
FAW Toyota	GAC Toyota	0.5689655172413790	33
Dongfeng Honda	GAC Honda	0.5625	27
Citroen	Peugeot	0.5595238095238100	47
SAIC GM	Shanghai GM	0.53125	34
Hyundai	Kia	0.45794392523364500	49
Beijing Hyundai	Dongfeng Yueda Kia	0.425531914893617	20
Opel	Peugeot	0.41304347826087000	38
Porsche	Skoda	0.410958904109589	30
Mahindra & Mahindra	Tata	0.39285714285714300	22
Bentley	Lamborghini	0.38461538461538500	10
Audi	Porsche	0.3805309734513270	43
Fiat	Peugeot	0.375	30
Audi	Volkswagen	0.37423312883435600	61

Figure 26: Top 13 most similar auto makers by global Jaccard similarity, highlighting two independent auto makers

We also observe that pairs lacking direct corporate or joint-venture ties can still exhibit considerable overlap if they share domestic ecosystems or operate in the same region. Mahindra and Mahindra & Tata, for example, record a Jaccard similarity of 0.393 with an intersection of 22 suppliers. Both are major Indian OEMs, and when looking at their list of shared supplier, we find that a large portion of their overlap stems from sourcing from local manufacturing infrastructure. Although this score remains below those of joint-venture pairs, it is still noteworthy for two fully independent automakers.

This shows how geographic proximity or domestic markets can drive partial supplier overlap, and corporate linkages, like a shared parent, tend to push similarities scores even higher.

From a strategic perspective, such partial overlap can shape competition in India's automotive market and may create opportunities for cost-sharing arrangements. On the other hand, it also raises the risk that a localized disruption will impact both OEMs simultaneously.

Product-Level Jaccard Analysis

Beyond the global Jaccard measures, when looking at product level Jaccard similarity across all pairs of makers, we also see some interesting results.

Namely,as shown in the red outlined in Figure 27 we observed that BYD, Brilliance, Donfeng Passenger Vehicle, FAW CAR, SAIC PAssenger Vehicle, all register a perfect similarity of 1 regarding their making of the power steering pump component. This indicates that these five major Chinese automakers all source from exactly the same set of nine suppliers for this critical component. Such perfect overlap strongly suggests these manufacturers are using standardized power steering pump designs, likely adhering to common technical specifications or industry standards.

This indicates creating a potential systemic vulnerability affecting these common suppliers would simultaneously impact multiple major automakers in the Chinese market. Additionally, the identical supplier base suggests limited product differentiation in this component across these domestic Chinese brands, especially when these makers are not all part of the same joint venture or parent company.

Maker1	Maker2	Product	jaccardSimilarity	intersection
BYD	Brilliance	power_steering_pump	1.0	9
BYD	Dongfeng Passenger Vehicle	power_steering_pump	1.0	9
BYD	FAW Car	power_steering_pump	1.0	9
BYD	SAIC Passenger Vehicle	power_steering_pump	1.0	9
Brilliance	Dongfeng Passenger Vehicle	power_steering_pump	1.0	9
Brilliance	FAW Car	power_steering_pump	1.0	9
Brilliance	SAIC Passenger Vehicle	power_steering_pump	1.0	9
Dongfeng Passenger Vehicle	FAW Car	power_steering_pump	1.0	9
Dongfeng Passenger Vehicle	SAIC Passenger Vehicle	power_steering_pump	1.0	9
FAW Car	SAIC Passenger Vehicle	power_steering_pump	1.0	9
Mazda	Mitsubishi	power_steering_pump	1.0	6
BYD	FAW Haima	power_steering_pump	0.8888888888888890	8
Brilliance	FAW Haima	power_steering_pump	0.8888888888888890	8
Dongfeng Passenger Vehicle	FAW Haima	power_steering_pump	0.8888888888888890	8
FAW Car	FAW Haima	power_steering_pump	0.8888888888888890	8
FAW Haima	SAIC Passenger Vehicle	power_steering_pump	0.8888888888888890	8
Fiat	Renault	starter_motor	0.8571428571428570	6
FAW Toyota	GAC Toyota	abs_esc	0.7777777777777780	7
Mazda	Nissan	torque_converter	0.7777777777777780	7

Figure 27: Product-level Jaccard similarity sorted in descending order. Red outline shows those who have exact same overlap. Blue outline shows the jaccard similarity between two joint ventures of Toyota.

Moreover, joint ventures with Toyota (for instance, FAW Toyota and GAC Toyota), as outlined in blue in Figure 27 exhibit high but not perfect overlaps for abs_esc systems, registering approximately 0.78.

The minor supplier diversification could serve multiple purposes. First, it may represent risk mitigation through slight supplier redundancy, allowing Toyota to maintain production if one supplier faces difficulties. Second, it could also reflect different regional requirements between northern China (FAW's base) and southern China (GAC's region). Third, the different vehicle segments targeted by each joint venture might necessitate slightly different ABS/ESC specifications despite sharing the same core technology.

Weighted Jaccard Analysis

From our results, we observe that Nissan and Toyota are two makers in the dataset with the greatest number of shared components (n=7). When running the weighted Jaccard on these two makers, with weights that prioritize safety-critical importance we produce the following intermediate table in Figure 28. (Implementation details in C.4)

	Maker1	Maker2	Product	jaccardSimilarity	intersection	sharedSuppliers	weight	weighted_jaccard
51	Nissan	Toyota	steering_gear	0.411765	7	[JTEKT Corporation, Rongtai Industrial Develop...	2.00	0.823529
58	Nissan	Toyota	stablizer	0.400000	8	[New Mather Metals, Inc., NHK of America Suspe...	0.70	0.280000
63	Nissan	Toyota	exhaust_manifold	0.388889	7	[Futaba Industrial Co., Ltd., Sango Co., Ltd.,...	0.05	0.019444
74	Nissan	Toyota	abs_esc	0.350000	7	[Bosch (Robert Bosch LLC), Aisin Corporation (...	0.90	0.315000
76	Nissan	Toyota	disc_brake_(pad)	0.344828	10	[Akebono Brake Corporation, ADVICS Co., Ltd., ...	2.00	0.689655
96	Nissan	Toyota	clutch	0.300000	6	[ZF Friedrichshafen AG, Schaeffler AG, P.T. EX...	0.20	0.060000
103	Nissan	Toyota	shock_absorber	0.285714	10	[KYB Americas Corporation - Indiana, thyssenkr...	0.60	0.171429

Figure 28: DataFrame showing similarity between Nissan and Toyota. They share a total of 7 components. Weights are modeled after safety importance in a car

Each row in the table contains maker pairs, a product label, the unweighted Jaccard, and a final weighted contribution. For instance, steering_gear carries a weight of 2.00, while exhaust_manifold carry only 0.05 because the former is critical to driver control while the latter has minimal impact on safety outcomes.

Their raw Jaccard sum is around 0.35, whereas the weighted total reaches 0.37. This seemingly small difference masks significant insights - safety-critical components show much stronger convergence than the overall supply network would suggest. The steering gear especially jumps out with an extremely high weighted Jaccard of around 0.82, revealing that these ostensible competitors have nearly identical supplier networks for this vital component.

For risk management teams, these findings paint a concerning picture: disruptions affecting key steering and braking suppliers would cascade through both manufacturers simultaneously, potentially crippling production across two of Japan's automotive giants. Yet the analysis also reveals potential resilience strategies - both manufacturers maintain distinctly different supplier relationships for shock absorbers and clutch systems, suggesting possible pathways for diversification in sourcing strategies for other parts.

Weighted similarities thus cut through the noise of raw supplier overlap data, showing where automakers are truly competing or collaborating in parts that matter most relative to their goals.

4 Conclusion

4.1 Supply chain overview:

- Extract top suppliers per product using PostgreSQL
- Visualize dominant suppliers and key supply chain shift using Neo4j

We provide a comprehensive view of the automotive supply chain by identifying top suppliers using PostgreSQL and visualizing key shifts with Neo4j.

4.2 Time series analysis:

- Analyze market share shifts, YoY changes using PostgreSQL
- Conduct statistical analysis in automaker partnerships, component competition, dominant suppliers using PostgreSQL
- Dive into Bosch & ZF to inspect maker regional distribution, product focus using PostgreSQL

For time series analysis, we tracked market share changes, automaker partnerships, and component competition. We also examined Bosch and ZF's regional distribution and product focus.

4.3 Maker's similarity:

- Find the most similar makers based on supply chain using Neo4j
- Weighted maker similarity analysis using Python

We analyzed maker similarity, using Neo4j to identify similar automakers and Python to add flexibility in weighted similarity analysis.

By uncovering both structural relationships and quantitative market trends, we answer diverse questions and offer deeper insights into supply chain dynamics and industry shifts.

5 Lessons Learned

Throughout this analysis of the automotive supply chain, several key insights and challenges emerged:

5.1 Analysis based

- **Supplier Network Complexity**: The complexity of supplier relationships varies across automakers, making it necessary to develop robust methodologies for quantifying similarity and identifying strategic supplier dependencies.

- **Evolving Supply Chain Dynamics**: The automotive supply chain is highly dynamic, with suppliers' roles and relationships shifting over time. Time series analysis provided valuable insights into how these changes occur and highlighted the importance of continuously updating datasets for accurate assessments.
- **Maker Similarity**: The similarity between joint ventures (like FAW Toyota and GAC Toyota) reveals how foreign partners create supply chain consistency across different regional partnerships while still allowing for measured adaptation to local conditions, resulting in high but imperfect similarity scores (0.78).

5.1.1 Tool based

- **Neo4j**: Leveraging the APOC library enabled efficient graph-based analysis of supplier networks.
- **PostgreSQL**: Used for data cleaning and text regularization to standardize information across different sources.
- **Python/Pandas**: Used for data pre-processing and useful for taking in outputs of Neo4j queries as CSVs to do further analysis in a pipeline.
- **Cross-Database Analysis**: Integrated multiple databases to enhance the depth and accuracy of the analysis.

This project provided valuable experience in researching a dynamic industry while also equipping us with practical skills in database management, data analysis, and supply chain modeling that will be applicable in future studies and industry applications.

6 Future Works

Having addressed the three key questions regarding the automotive supply chain, the next step is to analyze the impact of tariffs on the existing supply network. Future research could focus on the following areas:

- **Incorporating Tariff and Regulatory Data**: Integrating tariff rates and trade regulations into the analysis will provide a more comprehensive assessment of their impact on supply chain dynamics and supplier relationships.
- **Granular Product Analysis**: Refining product classifications by breaking them down into more specific products and gathering data on their suppliers will offer a more detailed representation of the supply chain structure. This will help identify key competitors in the automotive market more effectively and enhance the accuracy of supply network analysis.

References

[1] M. Race, "Six things that could get more expensive for americans under trump tariffs," 2025. [Online]. Available: https://www.bbc.com/news/articles/cvgpvq20qmdo

[2] "Marklines automotive data and analysis." [Online]. Available: https://www.marklines.com/en/

Appendix

A Data Preprocessing

A.1 Creating Master Table

We use code in Figure 29 to combine all the individual CSVs into one and generate two additional columns.

```python
# get all the csv files in the current directory
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# get all the csv files in the current directory
# do this in ascending order
# make sure the csv files are in this correct directory (same as this file), or edit the code to change argument to os.listdir)
csv_files = [f for f in os.listdir(".") if f.endswith(".csv")]
csv_files.sort()


# combine all the csv files
def read_csv_file(file_name):
    df = pd.read_csv(file_name, encoding_errors="replace")
    # add a column to the dataframe with the file name
    df["product"] = file_name[:-4].lower()  # remove .csv
    return df


# concat row wise
df = pd.concat([read_csv_file(file) for file in csv_files], axis=0)

df = df.drop_duplicates()

# Create a new column called trade_volume and have it be randomized uniformly between 100 and 1000, should be integer
df["trade_volume"] = np.random.randint(100, 1000, len(df))

# save the combined csv file
df.to_csv("combined_preprocessed.csv", index=False)
```

Figure 29: Code for merging individual components relational data into one, also generate two additional columns outlinine product type and dummy trading volume

B Time Series Analysis

B.1 Query: Supplier Collaborations by Region and Product

```sql
select distinct parent_company, region, model_region, product from component_fact c
left join supplier s on c.supplier = s.name
where  parent_company in ('Faurecia','FinDreams') and model_year = 2025
order by parent_company;
```

Figure 30: Result: Supplier Collaborations by Region and Product

B.2 Query: Market Share Trends of 2025's Top 10 Suppliers Over Time

```sql
with s_rank as (
select
    model_year,
    coalesce(s.parent_company,s.name) as supplier_name,
    rank() over(partition by model_year order by sum(unit) desc) as supplier_rank,
    sum(unit) as unit
from component_fact c
left join supplier s on c.supplier = s.name
group by model_year,supplier_name),
s_share as (
select
    r1.model_year,
    r1.supplier_name,
    r1.supplier_rank,
    r1.unit,
    sum(r1.unit) over(partition by r1.model_year) as unit_year,
    concat(round(cast(r1.unit / sum(r1.unit) over(partition by r1.model_year)*100 as numeric),2),'%') as share,
    case
        when r1.model_year = 2025 then concat(round(cast((r1.unit*12/3 - r2.unit) / r2.unit as numeric)*100,2),'%')
    else concat(round(cast((r1.unit - r2.unit) / r2.unit as numeric)*100,2),'%')
    end as yoy
from s_rank r1
left join s_rank r2 on r1.model_year = r2.model_year+1 and r1.supplier_name = r2.supplier_name)
select model_year, supplier_name, supplier_rank,share ,
       case
           when yoy='%' then null
           else yoy
       end as yoy
from s_share
where
    supplier_name in
     (select distinct supplier_name from s_rank
        where supplier_rank <= 10 and model_year = 2025)
order by model_year,supplier_rank;
```

Figure 31: Query:Market Share Trends of 2025's Top 10 Suppliers Over Time

B.3 Query: Case Study for Bosch & ZF

```sql
select model_year,c.region, maker_group,product,
       case
           when cn.region in ('W. Europe','C&E. Europe') then 'Europe'
           when cn.region in ('N. America','S. America') then 'Americas'
           when cn.region in ('India','South Korea','ASEAN') then 'ASEAN, India, Korea'
           when cn.region is not null then cn.region
       else model_region end as model_region,unit,
       count(*)over(partition by model_year) as count,
       sum(unit)over(partition by model_year) as yearly_unit,
       unit / sum(unit)over(partition by model_year) as share
from component_fact c
left join supplier s on c.supplier = s.name
left join maker m on c.maker = m.name and c.model_year = m.year
left join country cn on c.model_region = cn.nicename
where parent_company in ('Bosch','ZF')
group by model_year,c.region, maker_group, product,
         case
           when cn.region in ('W. Europe','C&E. Europe') then 'Europe'
           when cn.region in ('N. America','S. America') then 'Americas'
           when cn.region in ('India','South Korea','ASEAN') then 'ASEAN, India, Korea'
           when cn.region is not null then cn.region
       else model_region end ,unit
order by model_year, product;
```

Figure 32: Query: Case Study for Bosch & ZF

B.4 Query: Top 3 Suppliers in Each Product Category

```
with p_rank as (
select
    model_year,
    coalesce(s.parent_company,s.name) as supplier_name,
    product,
    rank() over(partition by model_year,product order by sum(unit) desc) as supplier_rank,
    sum(unit) as unit
from component_fact c
left join supplier s on c.supplier = s.name
group by model_year,supplier_name, product
order by model_year desc, product, supplier_rank)
select supplier_name, count(*) as high_rank_count,
       array_agg(model_year ||':'|| product order by model_year) as high_rank_products
from (select distinct supplier_name, model_year, product from p_rank where supplier_rank<=3) p
group by supplier_name
order by high_rank_count desc;
```

Figure 33: Query: Top 3 Suppliers in Each Product Category

C Maker Similarity

Note that since the database is extremely large (there are 16941 nodes and 49993 edges), we recommend running increasing the server.memory and dbms.memory settings in Neo4j in order to run the query. Specifically, we've used the following non-default parameters when running a Neo4j browser.

- `server.memory.heap.initial_size=16G`
- `server.memory.heap.max_size=16G`
- `server.memory.pagecache.size=16G`
- `dbms.memory.transaction.total.max=12G`

C.1 Turning Relational to Graphical Data through Neo4j

```
``` Cypher
// Load CSV with headers
LOAD CSV WITH HEADERS FROM 'file:///combined_preprocessed.csv' AS row

// Create unique nodes (using MERGE to prevent duplicates)
MERGE (region:Region {name: row.region})
MERGE (maker:Maker {name: row.maker})
MERGE (model:Model {
    name: row.model,
    model_year: toInteger(row.model_year),
    model_region: row.model_region
})
MERGE (supplier:Supplier {name: row.supplier})
MERGE (product:Product {name: row.product})

// Create relationships
MERGE (maker)-[:MAKES]->(model)
MERGE (maker)-[:LOCATED_IN]->(region)
MERGE (supplier)-[s:SUPPLIES]->(product)
MERGE (model)-[u:USES]->(product)

// Set the yearly volume as a property on the USES relationship since it
connects model and product
SET u.yearly_volume = toInteger(row.yearly_volume)
```
```

Figure 34: Script for loading in CSV to Neo4j

We use the script in Figure 34 to load in data mentioned in 2.2 to Neo4j to conduct analysis. There are three type of nodes and several types of edges.

C.2 Calculating Global Jaccard

To calculate the global Jaccard score for every maker, we use Neo4j to look for two distinct makers, then calculate their respective sets of suppliers across all models and all parts. This is shown in Figure 35.

```
Cypher
MATCH (m1:Maker)-[:MAKES]->(model1:Model)<-[s1:SUPPLIES_PART]-
(supplier:Supplier)-[s2:SUPPLIES_PART]->(model2:Model)<-[:MAKES]-(m2:Maker)
WHERE m1 <> m2 AND m1 < m2
WITH m1, m2,
     COUNT(DISTINCT supplier) AS intersection,
     COLLECT(DISTINCT supplier) AS sharedSuppliers
WITH m1, m2, intersection, sharedSuppliers,
    [(m1)-[:MAKES]->(model1:Model)<-[:SUPPLIES_PART]-(supplier:Supplier) |
supplier] AS m1Suppliers,
    [(m2)-[:MAKES]->(model2:Model)<-[:SUPPLIES_PART]-(supplier:Supplier) |
supplier] AS m2Suppliers
WITH m1, m2, intersection, sharedSuppliers,
     apoc.coll.union(m1Suppliers, m2Suppliers) AS unionSuppliers
WITH m1, m2, intersection,  unionSuppliers, sharedSuppliers,
     toFloat(intersection) / SIZE(unionSuppliers) AS jaccardSimilarity
WHERE intersection > 5
ORDER BY jaccardSimilarity DESC, m1.name, m2.name
RETURN m1.name AS Maker1, m2.name AS Maker2, jaccardSimilarity, intersection, [s
IN sharedSuppliers | s.name] AS sharedSuppliers
```

Figure 35: Script for calculating global Jaccard for every maker

C.3 Calculating Product-level Jaccard

To calculate the product Jaccard score for every maker, we use Neo4j to look for two distinct makers, condition on the same product, then calculate their respective sets of suppliers across all models and all parts. This is shown in Figure 36.

```cypher
MATCH (m1:Maker)-[:MAKES]->(model1:Model)<-[s1:SUPPLIES_PART]-
(supplier:Supplier)-[s2:SUPPLIES_PART]->(model2:Model)<-[:MAKES]-(m2:Maker)
WHERE m1 <> m2 AND s1.product = s2.product AND m1 < m2
WITH m1, m2, s1.product AS Product,
     COUNT(DISTINCT supplier) AS intersection,
     COLLECT(DISTINCT supplier) AS sharedSuppliers
WITH m1, m2, Product, intersection, sharedSuppliers,
    [(m1)-[:MAKES]->(model1:Model)<-[s1:SUPPLIES_PART]-(supplier:Supplier) WHERE
s1.product = Product | supplier] AS m1Suppliers,
    [(m2)-[:MAKES]->(model2:Model)<-[s2:SUPPLIES_PART]-(supplier:Supplier) WHERE
s2.product = Product | supplier] AS m2Suppliers
WITH m1, m2, Product, intersection, sharedSuppliers,
     apoc.coll.union(m1Suppliers, m2Suppliers) AS unionSuppliers
WITH m1, m2, Product, intersection, unionSuppliers, sharedSuppliers,
     toFloat(intersection) / SIZE(unionSuppliers) AS jaccardSimilarity
WHERE intersection > 5 // Filter by intersection size
ORDER BY jaccardSimilarity DESC, m1.name, m2.name, Product
RETURN m1.name AS Maker1, m2.name AS Maker2, Product, jaccardSimilarity,
intersection, [s IN sharedSuppliers | s.name] AS sharedSuppliers
```

Figure 36: Script for calculating product level Jaccard for every maker

C.4 Weighted Jaccard

To calculate the weighted Jaccard score between two makers, we first must download the outputs of the query for product level Jaccard mentioned in C.3. One can do this by running the query in the Neo4j browser and hitting download. After downloading the file as a CSV, we can feed this directly into a Python environment with Pandas installed to run the script shown in Figure 37.

```python
df = pd.read_csv('model_supplier_product_remove_duplicates.csv')
product_weights = {
    'steering_gear': 2,
    'stablizer': 0.7,
    'exhaust_manifold': 0.05,
    'abs_esc': 0.9,
    'disc_brake_(pad)': 2,
    'clutch': 0.2,
    'shock_absorber': 0.6
}

def maker_weighted_sim(df, maker_1, maker_2, weights: dict[str: float]):
    df = df.copy()
    maker_1 = df['Maker1'] == maker_1
    maker_2 = df['Maker2'] == maker_2

    df = df[maker_1 & maker_2]
    df['weight'] = df['Product'].map(product_weights)
    df['weighted_jaccard'] = df['jaccardSimilarity'] * df['weight']
    df['weight'] = df['weight'].fillna(1.0)

    total_weighted_sum = df['weighted_jaccard'].sum()
    total_weights = df['weight'].sum()

    normalized_score = total_weighted_sum / total_weights
    return normalized_score
```

Figure 37: Script for calculating the weighted Jaccard similarity between two makers

Note that the function takes in the dataframe that is loaded in, and takes in the name of the two makers, along with the dictionary of weights mapping product (as strings) to a specific weight. A default weight of 1 is applied to any product that is not included in weights dictionary. The output is then the normalized weighted similarity (between 0 and 1).