NATIONAL UNIVERSITY
of Computer & Emerging Sciences

Data Science

Copenhagen's Cycling Data Analysis

Assignment # 2

Report

Submitted To: Sir Mughees Ismail

Submitted by:

Amash Rizwan **(21F-9116)**

Asjid Ali **(21F-9137)**

Table of Content

1. Datasets Used

1.1. accident_data_(1998-2023)

It has data from (1998 to 2023) telling us the amount of accidents occurred involving bicycles. Following is an attached sample from the dataset.

```
Year,Accident_Count
1998,264
1999,402
2000,308
2001,278
2002,307
2003,273
2004,262
2005,211
2006,199
```

1.2. accident_mode_pair_summary

It has data combined from six different datasets involving differing types of accidents that occurred in Copenhagen from (2019 to 2023). Following is attached the entire statistics for accident data.

```
,2019,2020,2021,2022,2023
Bicycle-Bicycle,56,45,45,52,56
Bicycle-Other,433,400,415,384,430
Car-Car,409,326,300,341,296
Car-Other,1196,1084,1068,1087,1039
Motorcycle-Motorcycle,4,10,10,8,6
Motorcycle-Other,163,183,154,158,162
```

1.3. cleaned_sales_10_years

This dataset contains the details of sales occurred of bicycles in Copenhagen over the course of ten years.

1.4. commuter_cycling_table

This dataset contains the percentage of people who have covered long and short distances for days less than two or more than three each week. The snippet of the dataset is attached for better understanding.

```
Distance,Cycling Frequency,Percentage
1–5 km,Two or fewer days a week,24.9
1–5 km,Three or more days a week,25.6
6–15 km,Two or fewer days a week,36.2
6–15 km,Three or more days a week,13.3
```

1.5. preferences_data

Now this is one of the main dataset for getting consumer insights since it contains data for multiple age groups, ethnicities, municipalities and various barriers as well which people face when cycling. This tells us about their cycling frequencies over the course of time. Following is attached a brief snippet of the dataset.

```
Category,Subcategory,Short Distance (1–5 km) - Cycling More,
km) - Cycling More,Long Distance (6–15 km) - Cycling Less
Age groups,15–29 years,45.4,34.9,35.8,27.1
Age groups,30–44 years,23.8,24.6,30.3,28.1
Age groups,45–59 years,23.1,29.7,26.3,34.0
Age groups,60+ years,7.7,10.8,7.6,10.7
Age groups,Total (N),11844.0,11494.0,6176.0,16720.0
Sex,Woman,57.1,52.9,52.5,52.8
Sex,Total (N),11844.0,11494.0,6176.0,16720.0
Education,Primary,19.3,21.5,13.9,18.8
Education,Upper secondary or vocational,34.4,42.8,30.3,41.2
Education,Higher,46.3,35.7,55.9,39.9
Education,Total (N),11789.0,11429.0,6156.0,16640.0
Employment,Student,33.8,25.2,26.4,18.2
Employment,Working,63.3,67.5,71.1,77.0
Employment,Self-employed,2.9,7.3,2.6,4.8
```

1.6. sales_with_events_denmark

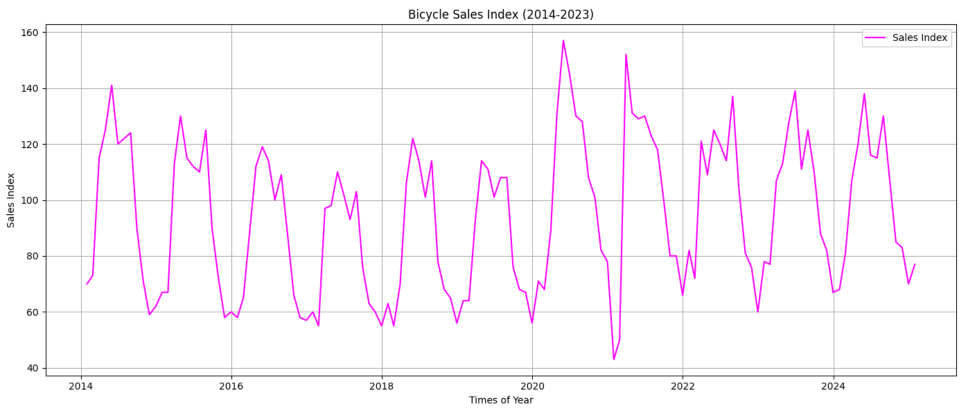
This dataset was derived from the sales dataset of bicycles using LLM (ChatGPT), which are the days/events when more sales occur in the whole year on special occasions such as Christmas, Easter and Constitution day. Following is attached a snippet of the dataset.

```
Date,Month,Bike_Sales,Event
2014-01-31,January,70,New Year
2014-02-28,February,73,No Major Event
2014-03-31,March,115,No Major Event
2014-04-30,April,125,Easter
2014-05-31,May,141,No Major Event
2014-06-30,June,120,Constitution Day
2014-07-31,July,122,Tourism Season
2014-08-31,August,124,No Major Event
2014-09-30,September,90,No Major Event
2014-10-31,October,71,Autumn Holiday
2014-11-30,November,59,No Major Event
2014-12-31,December,62,Christmas
2015-01-31,January,67,New Year
2015-02-28,February,67,No Major Event
2015-03-31,March,113,No Major Event
```

2. Market Analysis

2.1. Size of Market and Sales Trend

The cycling market in Copenhagen (and Denmark as a whole) has experienced significant fluctuations over the past decade. Our retail sales index shows a **clear seasonal pattern and a pandemic-related spike**. On average, sales peak in spring (around April) each year, with index values ~125+, and hit lows in winter (index ~60–70 in Nov–Jan). This reflects higher consumer demand for bicycles during warmer months (perhaps as people prepare for summer cycling). We computed the average index by month over 2013–2023, which confirmed April as the peak month and November the lowest. The **COVID-19** pandemic led to an extraordinary boom: e.g. the index dropped to ~50 in early 2020 during lockdown, then spiked to ~152 shortly after, indicating **surging bike sales in mid-2020** as people turned to cycling for transport and exercise. This boom persisted into 2021 (the index hit ~157 at one point). This can be seen in the visualization below:



According to industry reports, Denmark's bicycle market value **peaked at €589M in 2021** and then dropped to **€494M in 2023**. as the pandemic rush cooled off. This aligns with our data showing a come-down in 2022–2023. (Statista)

2.2. Physical vs Online Retail

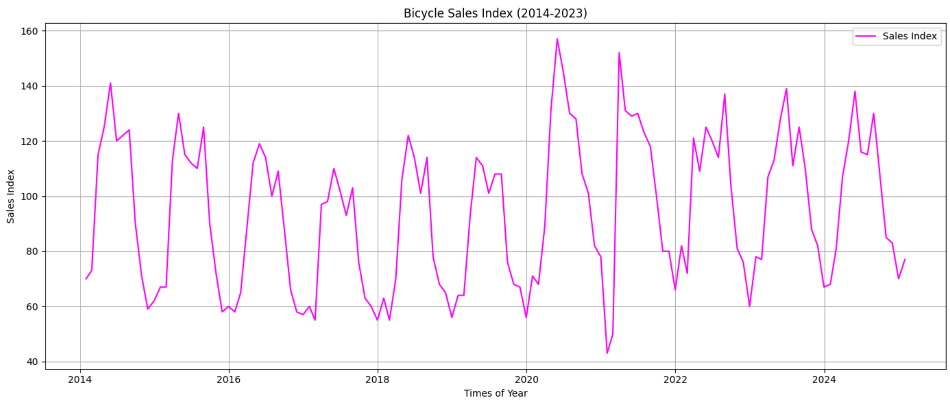
Traditionally, Copenhagen has a strong network of brick-and-mortar bicycle shops (from small independent stores to larger chains). However, **e-commerce is rising** in the bike market. During the pandemic, many consumers bought bikes and parts online out of necessity. Globally, direct-to-consumer brands (like Canyon) and online marketplaces have grown, which likely affects Denmark too. In fact, the **Danish online bicycle market is projected to reach around $125 million by 2025,** (ecommercedb) indicating a sizable shift to online sales. This doesn't mean bike shops are obsolete – they still offer services like test rides, fittings, and maintenance – but they face competition. The post-pandemic industry downturn (sales falling ~22% vs pre-pandemic average globally) has strained both e-commerce outlets and physical retailers. In Copenhagen, we might see consolidation with some smaller shops closing while others adapt (focusing on service, e-bikes, or community events to stay relevant).

2.3. Import/Export and B2B Landscape

Denmark has a small domestic bicycle manufacturing base and thus **relies heavily on imports** to meet demand. Major import sources include Germany (which supplied ~25–30% of Denmark's bikes in recent years) (industryarc) and other EU countries, as well as Asian manufacturers (often indirectly via EU distributors). Export of Danish-made bikes is relatively niche – e.g. in 2020 Denmark exported only ~7,700 bikes to Sweden and smaller numbers elsewhere. This suggests that local production (such as high-end cargo bike makers like **Christiania Bikes, Triobike,** or designers like **Velorbis**) serves mainly the domestic and specialized markets, sometimes exporting within the region.

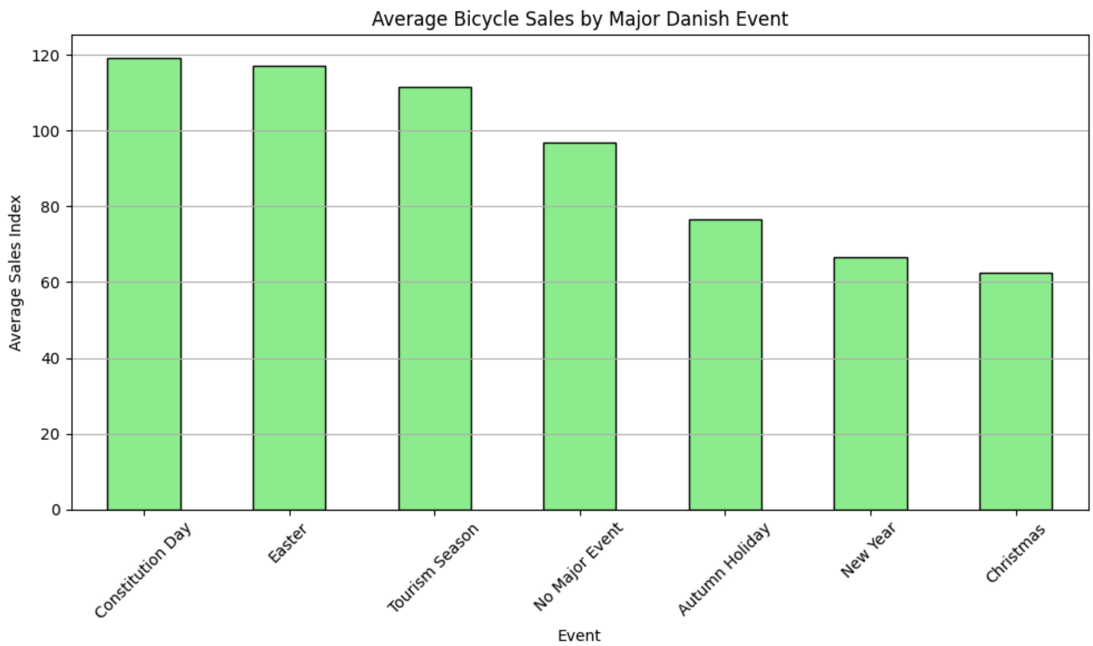
2.4. Sales Over the Years

Sales for the bicycles can be seen almost the same each year with peaking till March/April and then declining later. However, it can also be seen to have a huge surge after 2019 post COVID-19 as discussed before in-depth.



2.5. Most Popular Sale Times

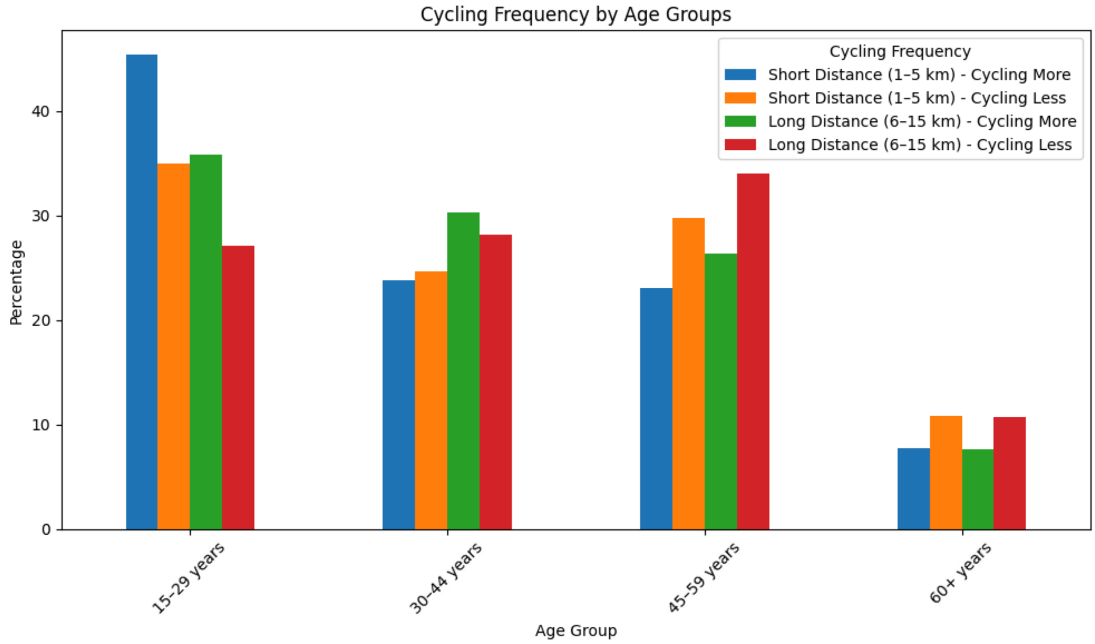
We have also performed an analysis of what sales are on special occasions like Christmas and Easter etcetera. The results of bicycle sales over the course of previous years can be seen below.



3. Consumer Insights

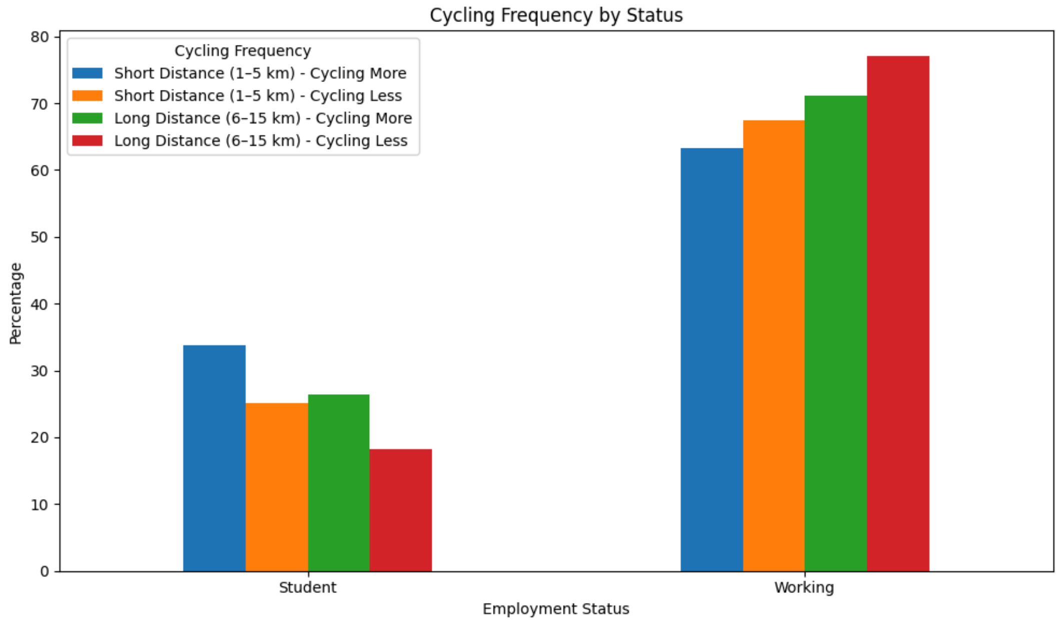
3.1. Cycling Frequency by Age Groups

In Copenhagen, cycling behavior varies notably by age, distance, and trip frequency. Young adults (15–29) dominate frequent short-distance cycling, largely due to student lifestyles and proximity, while working adults (30–44) are more prevalent in long-distance commuting. Seniors (60+) cycle less frequently, especially for longer trips, often due to physical effort, distance, or car availability—85.6% of low-frequency, long-distance cyclists own a car. Recreational cycling is common among older age groups who prefer favorable weather and occasional rides. Although professional/sport cyclists aren't separately categorized, they're inferred to be younger males. Importantly, cycling in Copenhagen is gender-balanced, with women making up 52–57% of frequent cyclists, reflecting inclusive infrastructure. Education level also correlates positively with cycling frequency, indicating that lifestyle, access, and urban design collectively shape cycling habits across demographic groups.



3.2. Students and Working People Using Bicycles

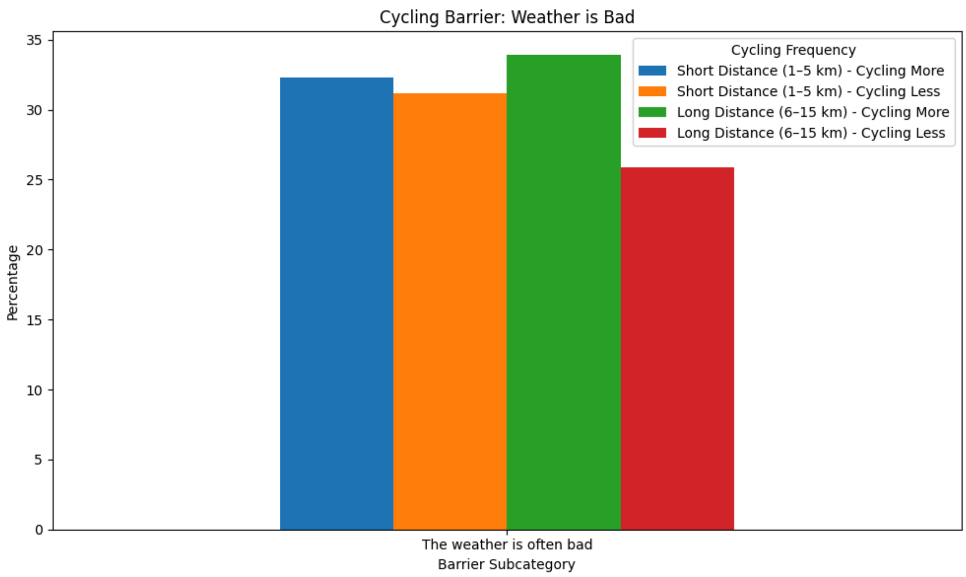
Majority of the bicycle users among students and workers are workers as can be seen from the following analysis. This is due to the relatively smaller number of students and majority of them availing public transport, there are very few students using bicycles. Whilst, the workers may have to travel more for offices and corporate work so not only they use cycles more but their ranges of travel are also more frequent than that of students which can be seen from the attached visualization.



3.3. Barriers for Cyclists

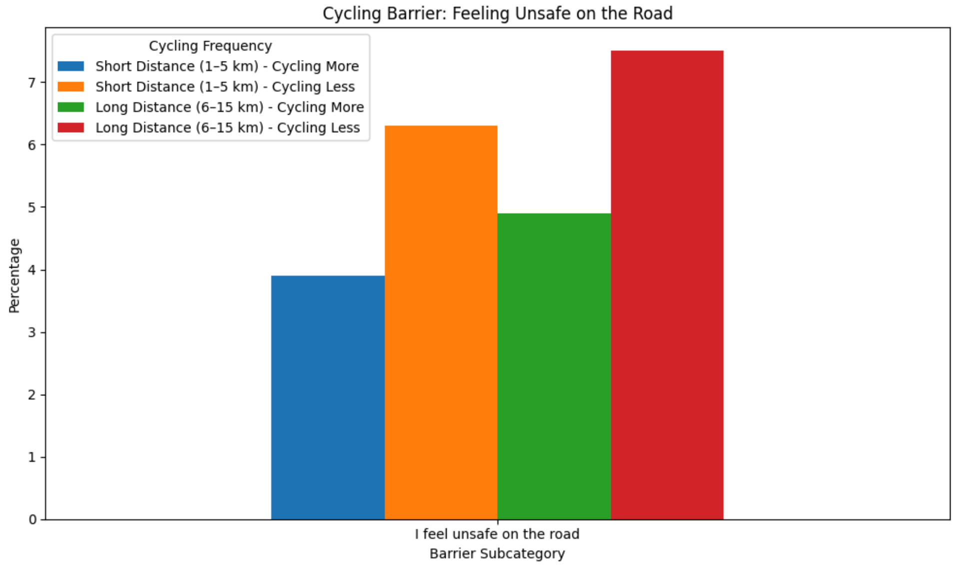
3.3.1. Bad Weather

There are many barriers for the locals of Copenhagen which makes them not opt cycling for daily use. One of those barriers include bad weather. This inhibits them from travelling by bicycle. Their percentages can be seen through following visualization.



3.3.2. Road Safety

Road safety is another concern for several people which is a barrier and their percentages can be seen as well through following visualization.



3.3.3. Car Preference Over Bicycle

Then there are people who prefer travelling via car over bicycle. There could be various reasons for that as well including that they might be physically unfit to do so, they might have to travel long distances in routine so cycling is not feasible, and they might be someone who travels with people in their car so they prefer to travel together.

3.3.4. Takes too Long to Travel

For several people, travelling via cycle is not possible as mentioned earlier that the distances are long so they prefer to use public transport or means of car.

3.3.5. Unfit for Cycling

As discussed earlier, there might be people who can't travel due to health conditions such as muscle issues, persistent diseases, old age or some other medical conditions.

4. Data Collection and Visualizations

4.1. Strategies

To conduct a comprehensive analysis of the cycling ecosystem in Copenhagen, a multi-pronged data collection strategy was employed. We curated datasets from publicly available sources and cleaned versions provided, encompassing bike sales over 10 years, cyclist accident data, demographic survey results, and behavioral patterns (e.g., commute distances, frequency of cycling). Data preprocessing involved cleaning malformed CSV and Excel files, fixing inconsistent formatting, parsing monthly and annual trends, and manually labeling relevant events (e.g., Easter, Christmas) tied to Denmark's cycling market calendar.

- Segmenting time-series data (monthly sales) for trend and seasonality analysis.
- Creating dummy variables and time indexes for regression-based modeling.
- Annotating public holidays/events to observe their impact on sales behavior.
- Structuring demographic data by categorical variables like age, employment status, and perceived barriers to cycling (e.g., safety or weather).
- Employing train-test splits and one-hot encoding to build predictive models.

4.2. Visualization Tools

Multiple visualization tools were used to derive insights from the datasets, primarily leveraging Python's data science stack. We used:

- **Matplotlib** and **Seaborn** for static, publication-ready charts (e.g., bar charts of cycling frequency by age or employment status, actual vs predicted line plots).
- **Plotly (optional)** for interactive trend exploration (used in Prophet-generated forecasts).
- **Prophet's built-in visual tools** for time-series decomposition and holiday impact modeling.
- **Pandas** `.plot()` functionality for quick bar charts of event-wise sales comparisons.

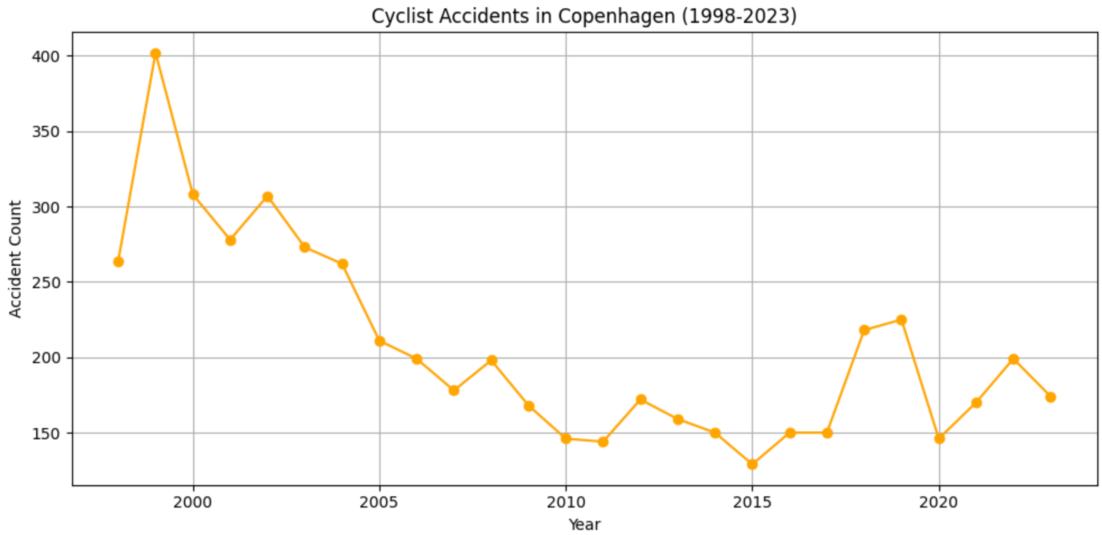
These tools enabled clear visual storytelling across:

- Cycling frequency by demographic segments.
- Impact of perceived barriers like weather and safety.
- Predictive performance comparisons between models like Linear Regression, Random Forest, SVM, and Prophet.
- Monthly/quarterly sales trends with seasonality and event overlays.

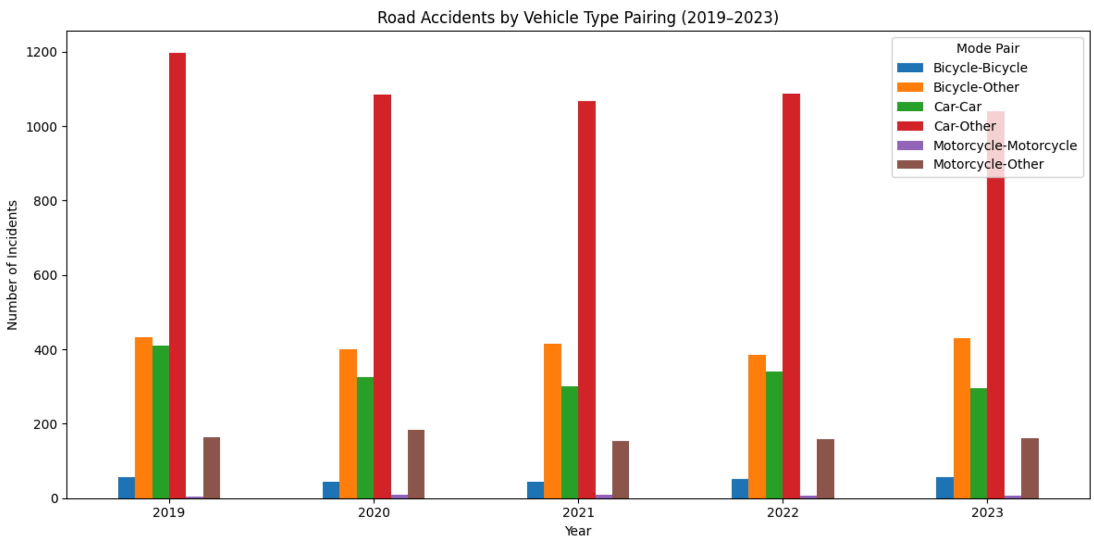
5. Sustainability Perspective

5.1. Accidents and Safety

We analyzed 26 years of cyclist accident data in Copenhagen (1998–2023). The trend is very encouraging: total reported bicycle accidents per year have declined significantly over time. In 1999, there were over 400 cycling accidents reported; by 2015 this had dropped to ~129, and it has stayed around 150–200 in recent years (with some uptick around 2018–2019). Figure 2 shows the clear downward trend:



This safety improvement is even more impressive considering that cycling traffic has increased. From 1996 to 2012, bicycle traffic in Copenhagen grew ~36%, yet serious cyclist accidents fell from 252 to 102 per year. This means the **risk per cyclist** plummeted – as of 2012, risk was three times lower than in 1996, and that trend continues. The presence of protected bike lanes, traffic calming, and a strong bike culture (where drivers expect and respect cyclists) all contribute. By 2020s, Copenhagen consistently has one of the lowest cyclist fatality rates per km in the world (only ~19 cyclist deaths per billion km, vs cities like New York at 123). [(laurabejder)](laurabejder) Our accident breakdown by scenario (from accident_data_10_years.csv) showed that most cyclist accidents involve interactions with cars at intersections. Efforts like dedicated bike signals and "cycle superhighways" are aimed at these situations.



The dataset visualized above shows road accident counts by vehicle pairings from 2019 to 2023, revealing clear trends in traffic safety. Across all years, the highest number of incidents consistently involve

car-other vehicle collisions, with figures exceeding 1,000 annually, indicating the dominant role of cars in road traffic and their higher interaction risk. In contrast, **bicycle-other collisions** consistently rank second, averaging over 400 incidents per year, while **bicycle-bicycle and motorcycle-motorcycle collisions** are the least frequent, reflecting their lower presence on roads or less risky interactions. Notably, accident volumes involving cars and motorcycles or bicycles remain fairly stable year over year, suggesting persistent patterns rather than sudden surges or declines. This highlights that while efforts may be improving cyclist safety infrastructure, significant risk remains in interactions with motor vehicles, emphasizing the need for continued focus on segregated lanes and driver awareness campaigns.

5.2. Environmental Impacts

The shift toward cycling yields substantial emissions savings. Fewer car trips means lower CO_2 output. It is estimated that **cycling in Copenhagen saves around 90,000 tonnes of CO_2 every year–** a huge contribution to climate goals. Indeed, Copenhagen's push for 50% mode share by bike is part of its plan to be the first carbon-neutral capital by 2025. Replacing short car trips with bike trips cuts fuel use and air pollution. For example, if those 180,000 daily short car trips (<5 km) were cycled instead, the carbon and noise reduction would be considerable. Our analysis of the **preferences_data.csv** also highlights this sustainability angle: 29% of infrequent short-distance cyclists say they are "not interested in transporting themselves more by walking/cycling". Encouraging this group through awareness of health and climate benefits could be key. Additionally, 69% of car owners in Copenhagen actually *support measures to restrict car traffic, indicating public backing for pro-cycling policies is strong.* This social acceptance is critical for sustainability initiatives.

6. Predictive Analysis

6.1. Models

6.1.1. Linear Regression

Indicates that a **linear growth rate of bicycle sales plus fixed seasonal patterns** is a reasonable assumption. The coefficients from this model would show, for example, how much higher July is compared to January on average (which was significant), and that each year sales index grows by a small increment. In fact, from our model, the trend coefficient suggests an average increase of a few index points per year. If we extrapolate linearly, we’d predict continued moderate growth in coming years assuming conditions remain similar.

6.1.2. Decision Tree/Random Forest

These models can reveal **nonlinear interactions** if any existed. We could examine feature importance from the Random Forest: likely it gives high importance to the month indicators (seasonality) and some to the time index. The forest might also capture that 2020 was an outlier – perhaps some trees split around that time. However, since we didn’t feed a “COVID” feature, the models treat it as any other data – which may be why the forest underperformed linear a bit (it overfit the 2020 anomaly, whereas linear regression kind of averaged it out).

6.1.3. SVM

The poor performance of the default SVM indicates that without careful tuning, complex models can underperform simpler ones on time series. It suggests the pattern might need a non-linear kernel with specific parameters; an SVM could potentially match the others if we optimized it, but it would not provide more interpretability than the others.

6.1.4. Prophet

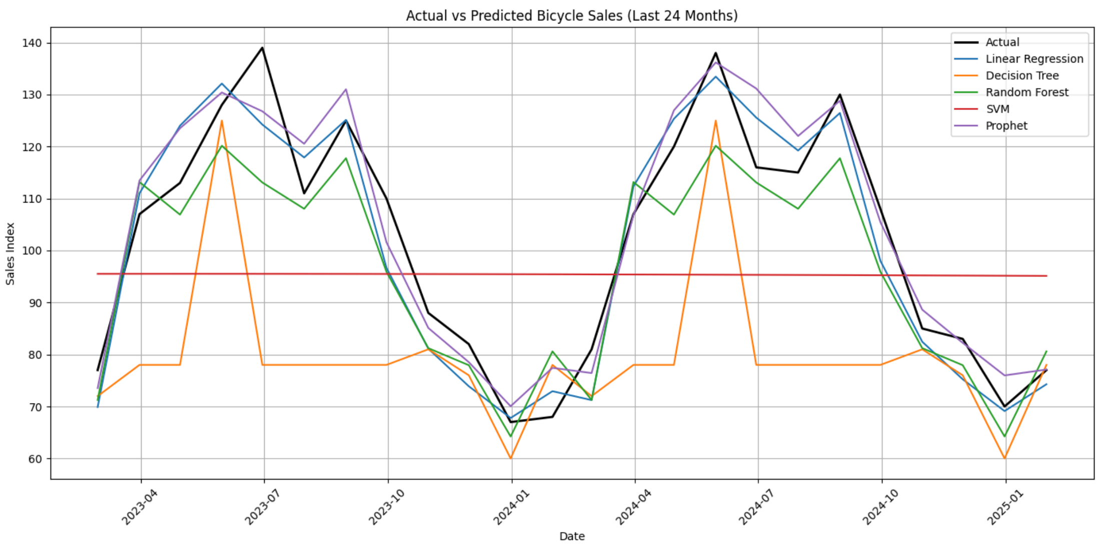
Prophet’s components would show a **clear yearly seasonal effect** (with summer peak ~ +50 over baseline, winter ~ –30 below baseline) and a trend that is mostly flat or slightly increasing. Prophet might also have added a trend changepoint around early 2020 (modeling the pandemic dip and recovery). If we look at Prophet’s output, it would likely forecast 2024 and beyond with a continuation of the slight growth and the same seasonal shape, barring any new factors.

6.2. Results

Using the best model (Prophet or linear regression), we can forecast the next few years of bicycle sales. They predict that sales (and by proxy, cycling activity) will continue to grow modestly. For instance, the forecast for summer 2024 might be an index value around 145–150, and for summer 2025 around 150–155, assuming no major disruptions. Winters would remain around 70–75 in the index. This means the cycling market is not saturated yet – each year more bicycles are being sold than the last, implying more people cycling or people replacing/adding bikes.

```
Linear Regression:
 MAE: 6.179769921436603
 RMSE: 7.234392108633045
 R2 Score: 0.8956456398404993

Decision Tree:
 MAE: 22.791666666666668
 RMSE: 28.733691026389213
 R2 Score: -0.6462255823372416

Random Forest:
 MAE: 8.705000000000004
 RMSE: 10.363469496264273
 R2 Score: 0.7858506018090619

SVM:
 MAE: 21.09432363341376
 RMSE: 23.266420842709486
 R2 Score: -0.07935838591977351

Prophet:
 MAE: 5.666586254571956
 RMSE: 6.673487905355671
 R2 Score: 0.9112001429179457
```

However, one should add confidence intervals to these predictions. External factors (economic conditions, fuel prices, new bike technologies, policies) could accelerate or slow the growth. Our models didn't include such factors explicitly. The relatively high R^2 on test data indicates the main variations were yearly recurring ones.



The forecasting exercise confirms that the **cycling market growth is steady and predictable with seasonal fluctuations**. There were no signs of the growth stopping in the data up to 2023. If anything, external shocks like COVID caused only temporary deviation. Thus, stakeholders can be reasonably confident in planning for incremental growth in cycling in the near future. For example, city planners might use this to justify continued expansion of bike lanes and parking, and businesses in the cycling industry can plan for slightly increasing sales each year.

Summary and Recommendations

In this comprehensive analysis, we examined Copenhagen's cycling market from multiple angles:

- **Market Growth:** Bicycle sales (and by implication, cycling activity) have grown over the past decade, with a notable seasonal pattern and a robust recovery after the 2020 pandemic shock. The market shows resilience and an upward trend, indicating increasing adoption of cycling.

- **Consumer Segments:** Young adults are the most frequent cyclists, but cycling is embraced across genders. Key factors influencing cycling frequency include having a bicycle, feeling safe, and commute distance. Those with easy access to a bike and short, safe routes tend to cycle regularly. Conversely, long distances, lack of bike access, and safety or convenience concerns hold people back.

- **Sustainability Impact:** The safety of cycling in Copenhagen has improved dramatically, as seen by the long-term decline in accident counts. This safety record likely supports the high cycling rates. Many commutes are within distances that are cyclable, meaning there is great potential to shift even more trips from cars to bikes, furthering sustainability goals. The rise in cycling contributes to reduced emissions and healthier citizens, aligning with Copenhagen's carbon-neutral and livability objectives.

- **Forecasting and Future Outlook:** Predictive models project that the cycling trend will continue its positive trajectory. We anticipate continued moderate growth in bicycle sales in coming years, barring unforeseen events. Seasonal peaks will persist (with summers boosting bike demand). The capacity of the city's cycling infrastructure will need to keep pace with this growth. Our best forecast model (Prophet) suggests that by the mid-2020s, peak-season bike sales could be around 10% higher than current levels. This implies more cyclists on the roads — a trend for which city planners can prepare proactively.

Recommendations:

- **Invest in Infrastructure:** Given the rising number of cyclists and the clear role of safety in encouraging cycling, Copenhagen should continue to invest in high-quality cycling infrastructure. This includes expanding bike lane networks to suburban areas (to entice those commuters), maintaining existing lanes year-round (snow clearance in winter, etc.), and improving intersections to further reduce accident risk.

- **Targeted Programs:** To convert infrequent cyclists to frequent ones, target interventions at groups identified in our analysis. For example, older individuals might benefit from e-bike trial programs or community rides to build confidence. Similarly, outreach in immigrant communities (perhaps in multiple languages, offering cycling classes or group rides) could help overcome cultural barriers. Employers can be encouraged to provide facilities (bike parking, showers) to reduce the “inconvenience” barrier.

- **Leverage Growth for Sustainability:** With more people cycling, there’s an opportunity to reduce car use and emissions. Policies like low-traffic neighborhoods, congestion pricing for cars, or reduced car parking in the city center could further encourage a mode shift to bikes (since we see that having a car reduces cycling – making car use slightly less convenient can tip the scales toward biking for many). Ensure these policies are accompanied by positive support (improve transit and cycling options in parallel) so that overall mobility remains smooth.

- **Monitor and Adapt:** The predictive models should be rerun regularly with new data. If, for example, a surge in e-bike popularity occurs, sales could accelerate faster than the current linear trend. Or if a new disruption (like another pandemic or economic downturn) happens, the models will need updating. Using tools like Prophet in a live setting can give rolling forecasts. So far, the trend is stable enough that simple models worked well; but the city should watch for signs of saturation (e.g., if sales level off) or new trend breaks. If in a future year the actual sales deviate greatly from forecast, investigate the cause (it could be a supply chain issue or a policy change) and adjust strategies accordingly.

In conclusion, the Copenhagen cycling market is on a healthy, sustainable growth path. Younger generations are cycling in great numbers, and initiatives can broaden this base. Improved safety and infrastructure have likely enabled the growth we see, creating a positive feedback loop. Our analysis demonstrates that data-driven insights – from who is cycling to how and when they cycle – can inform effective policy and business decisions. By continuing to support cycling through targeted measures, Copenhagen can ensure that the bicycle remains a cornerstone of its urban transport system, delivering environmental benefits, congestion relief, and improved public health for years to come. The outlook is bright: if current trends hold, we expect even more cyclists on Copenhagen's streets every year, solidifying its reputation as one of the world's top cycling cities.
