F21-9103

F21-9294

F21-9141

Cycling Market Data Collection & Visualization

**1. List of Data Sources with URLs**

**Data Sources:**

1. **Statistics Denmark** (Cycling and sustainability reports in Denmark):
   * **URL**: <https://www.dst.dk/en>
   * **Description**: Provides official reports and datasets on cycling infrastructure, urban transportation, and sustainability in Denmark.
2. **Yelp API** (Cycling shop reviews and ratings in Copenhagen):
   * **URL**: <https://www.yelp.com/developers>
   * **Description**: Offers real-time data on cycling shops, including customer reviews, ratings, and business information.
3. **ITC Trade Map** (Bicycle trade data for Denmark):
   * **URL**: <https://www.trademap.org/>
   * **Description**: Provides structured trade data on the import and export of bicycles in Denmark and other countries.
4. **DBA.dk + Cykelpartner.dk** (E-commerce platforms for bicycles):
   * **DBA.dk URL**: <https://www.dba.dk>
   * **Cykelpartner.dk URL**: <https://www.cykelpartner.dk>
   * **Description**: These platforms provide real-time data on bicycle sales, consumer trends, and pricing for new and used bicycles in Denmark.
5. **Bike Europe** (Cycling industry market research):
   * **URL**: <https://www.bike-eu.com/>
   * **Description**: Offers market insights and reports on cycling industry trends, growth, and forecasts in Europe.

**2. Detailed Explanation of Data Acquisition Methods**

**Data Acquisition Methods:**

1. **Statistics Denmark**:
   * **Method**: We will download relevant reports and datasets on cycling and sustainability directly from the official website.
   * **Example**: Like if we search for keywords like "cycling infrastructure" or "urban transport" and can download reports in **CSV** or **PDF** format.
2. **Yelp API**:
   * **Method**: We will access data on cycling shops (reviews, ratings, business info) via API calls.
   * **Example**: Can Use the Yelp API to query cycling shops in Copenhagen by business type and location.
   * **Tool**: We will be using a Python script with requests or a similar API client to collect the data.
3. **ITC Trade Map**:
   * **Method**: We will do the Sign up for access to the ITC Trade Map portal and can download trade reports related to bicycle imports/exports for Denmark.
   * **Example**: We can download annual bicycle trade data in **CSV** or **Excel** format.
4. **DBA.dk + Cykelpartner.dk**:
   * **Method**: We will be doing web scraping to collect data on product listings, pricing, and sales trends from both platforms.
   * **Example**: We can use **Selenium** or **BeautifulSoup** or **Scrapy** in Python to scrape product listings, including price, condition (new/used), and availability.
5. **Bike Europe**:
   * **Method**: We will download free reports and articles on market trends and forecasts from the Bike Europe website.
   * **Example**: We will access relevant reports and can download them for analysis.

**3. Data Storage Strategy**

**Storage Strategy:**

1. **Relational Databases (MySQL/PostgreSQL)**:
   * **Use Case**: We will be storing structured data like government reports from **Statistics Denmark** and trade data from **ITC Trade Map**.
   * **Example**: We will be creating tables for data like bike sales, cycling infrastructure metrics, import/export volumes, and CO₂ savings.
2. **NoSQL Databases (MongoDB)**:
   * **Use Case**: We will be storing semi-structured data such as business listings from **Yelp API** and e-commerce data from **DBA.dk** and **Cykelpartner.dk**.
   * **Example**: We will be creating collections (documents) where each record represents a business or product listing, including fields like business name, ratings, reviews, price, and availability.
3. **Cloud Storage (AWS S3)**:
   * **Use Case**: We will be storing large datasets, raw scraped data (e.g., HTML from DBA.dk or Cykelpartner.dk), and backups.
   * **Example**: We will be uploading raw files and snapshots to AWS S3 for scalability and easy access.

**4. Justification of Data Source Reliability and Completeness**

**Reliability and Completeness:**

1. **Statistics Denmark**:
   * **Justification**: This is an official government source that provides comprehensive and regularly updated reports on cycling infrastructure and sustainability, ensuring a high level of credibility.
2. **Yelp API**:
   * **Justification**: Yelp provides real-time, user-generated reviews and ratings, making it a reliable source for understanding customer satisfaction and business performance for cycling shops.
3. **ITC Trade Map**:
   * **Justification**: ITC Trade Map is a well-recognized platform for official trade data, updated regularly, making it highly reliable and accurate for analyzing import/export trends.
4. **DBA.dk + Cykelpartner.dk**:
   * **Justification**: These are popular e-commerce platforms in Denmark, providing up-to-date information on product listings, sales trends, and consumer behavior, making them reliable sources for market analysis.
5. **Bike Europe**:
   * **Justification**: Bike Europe is an industry-recognized source offering credible and well-researched reports on market trends and forecasts, ensuring a reliable and comprehensive view of the cycling industry.

**5. Outline of Predictive Models**

**Predictive Models:**

1. **Linear Regression:**

**Use Case**: Predict overall market growth by analyzing historical bicycle sales data.

**Why it’s suited**: Linear regression is simple and interpretable, making it ideal for identifying linear trends in market growth.

**Prediction Target:** Annual bicycle sales growth in Copenhagen.

1. **Facebook Prophet:**

**Use Case:** Time series forecasting for bicycle sales trends and seasonal variations.

**Why it’s suited:** Facebook Prophet is designed for time series data, excelling at handling seasonality and trend changes.

**Prediction Target:** Monthly or quarterly bike sales variations, including seasonal peaks.

1. **Random Forest:**

**Use Case:** Predict changes in consumer behavior based on multiple market indicators.

**Why it’s suited:** Random Forest reduces overfitting and handles complex, mixed data effectively.

**Prediction Target:** Consumer demand shifts across bike types (e.g., commuter vs. recreational).