

High-Level Proposal for Health Foods for Dogs Forecast Project

Objective

The primary objective of this project is to identify optimal health food options for dogs by analyzing historical sales data and social media reviews from platforms such as Instagram, Twitter, and pet-specific forums. Given the increasing trend among dog owners to choose organic and healthier food options for their pets, this project aims to provide actionable insights for businesses in the pet food industry to optimize their product offerings.

Background

There has been a significant shift towards healthier and organic food options for dogs, which often includes raw meat, essential vitamins, and other additives designed to promote better digestion and overall health. These diets are believed to extend a dog's lifespan, reduce the likelihood of cancer, enhance the immune system, and improve skin and dental health.

Businesses in the pet food industry, such as Petco, Rover, The Farmer's Dog, Chewy, Whole Foods, and BJ's Raw Meat, can leverage aggregated data to optimize their SKUs. Just as holistic medicine has influenced human diets, there is a growing movement to move away from processed kibble in favor of more natural food options for dogs.

Data Sources

1. **Sales Data:** Data from online pet food retailers and stores such as Amazon, Petco, and Whole Foods. Example sources include:
 - [Statista](#), [IBISWorld](#), [Future Market Insights](#)
2. **Social Media Data:** Data from APIs provided by Instagram, Twitter, and pet-specific forums. Influential accounts include:
 - [Kayla Kowalski's Guides](#), [Instagram: Pet Nutrition](#), [Instagram: Wellness Pet Food](#), [Instagram: Green Juju Kitchen](#)
3. **Additional Datasets:** Veterinary health data, nutritional information from pet food brands, and customer reviews from sites like Chewy.
 - [IMARC Group](#), [Statista](#), [Crawl Feeds](#), [Figshare](#)

Data Acquisition

1. **Sales Data:** Scrape data from pet food retailers or use available datasets on platforms like Kaggle.
2. **Health and Nutritional Data:** Collect data from veterinary and nutritional databases, as well as brand websites.
3. **Social Media Data:** Use APIs to collect posts, hashtags, and reviews related to dog health foods.
 - a. Action item: Check Rate Limiting in [Instagram API documentation](#)

Data Model

- **Products:** Different brands and types of dog food.
- **Reviews:** Social media posts, forum discussions, and customer reviews.

- **Nutritional Information:** Ingredients and health benefits.
- **Sales Transactions:** Historical sales data by product and time period.
- **Health Outcomes:** Data on the health impact of different foods on dogs.

Pipeline Design

1. **Data Extraction:**
 - Utilize APIs and web scraping to pull data from social media platforms and retail websites.
 - Extract nutritional and health data from veterinary databases and brand sites.
2. **Data Transformation:**
 - Clean and preprocess text data from reviews and posts.
 - Perform sentiment analysis to gauge public opinion on different dog foods.
 - Aggregate sales data by product categories and time periods.
 - Normalize nutritional information for comparison.
3. **Data Loading:**
 - Load processed data into a data warehouse such as AWS Redshift or Google BigQuery.
 - Ensure the pipeline can be easily updated to include new data sources or fields.

Resilience and Configurability

- Implement modular pipeline components that are easily modifiable.
- Use configuration files to manage data sources and processing logic.
- Set up logging and monitoring to ensure pipeline reliability and facilitate troubleshooting.

Query and Visualization

1. **Develop SQL Queries:**
 - Identify trends in dog food preferences.
 - Correlate sales data with social media sentiment.
 - Analyze the impact of specific nutritional ingredients on dog health.
2. **Create Dashboards:**
 - Use tools like Tableau or Power BI to visualize sales trends, popular products, and customer sentiment.
 - Provide actionable insights for pet food retailers and health-conscious pet owners.

Example Datasets

- **Sales Data:** Kaggle datasets for pet food sales.
- **Social Media Data:** Twitter API for hashtags like #dogfood, #healthydog, and Instagram posts from popular pet influencers.
- **Health Data:** Veterinary health databases and nutritional information from pet food brands.

Overall Takeaway

This project will provide a comprehensive skill set including data scraping, API integration, text analysis, and building robust data pipelines. By focusing on health foods for dogs, it addresses a niche yet growing market, offering practical benefits to pet owners and businesses alike.