Mô tả Dataset

1. Book Store

Category: Sales & Inventory

Date Added: 21-12-22

# of records: 669

# of Fields: 2676

The Book Store dataset provides information on book sales, ratings, and reviews. This data is valuable for book store owners, marketers, and researchers who are interested in understanding consumer behavior and preferences in the book industry.

By analyzing the dataset, book store owners can identify trends and patterns in consumer behavior, which can inform product selection and marketing strategies. Marketers can also use this data to develop targeted marketing campaigns that appeal to specific customer segments. Researchers can also use this data to understand consumer behavior and preferences in the publishing industry.

2. Call Center Logs

Category: Customer Service

Date Added: 01-07-22

# of records: 4000

# of Fields: 52000

This dataset contains information about customer complaints, including the type of complaint, the product or service involved, and the resolution or outcome. The data can be used to identify patterns and trends in customer complaints, and to develop strategies to improve customer satisfaction and retention.

By analyzing this dataset, businesses can identify areas where they need to improve their products or services, and can develop more effective complaint handling processes. This dataset is a valuable resource for businesses of all sizes, as it can help them to maintain a positive reputation and retain loyal customers.

3. Insurance Claims

Date Added: 01-07-22

# of records: 15258

# of Fields: 167838

This dataset contains information about insurance claims, including the type of claim, the claim amount, and the duration of the claims process. The data can be used to analyze patterns and trends in insurance claims and to develop insights into the factors that impact the claims process.

By studying this dataset, insurance companies and claims adjusters can gain insights into the factors that contribute to claim delays and inefficiencies and can develop effective strategies to improve the claims process and enhance customer satisfaction.

4. Commercial Airline Flights

Category: Transportation

Date Added: 01-07-22

# of records: 10001

# of Fields: 220022

The Flights dataset is a collection of data related to commercial airline flights. It includes information such as flight numbers, departure and arrival times, aircraft type, and flight routes.

This dataset can be used by airlines to optimize flight schedules, improve operational efficiency, and reduce costs. For example, an airline may use this data to identify routes

with low demand and adjust their schedules accordingly. Additionally, the dataset can be used by travel agencies and consumers to search for flights and compare prices across different airlines.

5. Telecom

Date Added: 20-12-22

# of records: 1980

# of Fields: 9005

The Telecom dataset provides information on customer usage patterns, service quality, and other factors related to telecommunications services. This data is valuable for telecom companies, regulators, and researchers who are interested in analyzing trends and patterns in the telecommunications industry.

By analyzing the dataset, telecom companies can identify areas for improvement in service quality, pricing, and customer satisfaction. Regulators can also use this data to develop policies that protect consumers and promote competition in the industry.

6. Environmental Data Reporting

You are tasked with developing a report to communicate relevant environmental data that helps visualize evolving patterns. The indicators have already been selected and the data has been collected.

Indicators:

- An index is a numerical value derived from actual measurements of a stressor, state, exposure, or human health or ecological condition over a specific geographic area, which trends over time represent or draw attention to underlying trends in environmental conditions. The report should include information on all four key environmental indicators for the area:
- Heavy rainfall: annual compilation of daily rainfall from 23 locations (22 airport locations, 1 non-airport location) throughout the region.

- Ambient air toxics: airborne toxic concentrations harmful to human health and ecology at locations in the region. Note that concentration is expressed in μg/m3 (micrograms per cubic meter of air).

- Asthma prevalence: adult asthma prevalence (via modeling) by census tract level and pediatric prevalence by state percentage across the district area. Note that in children's asthma data, "suppressed" means the number is too low to show without potentially revealing someone's identity and is "unstable" (often overlapping repetition with "restrained") means that the numbers used to calculate the rate are so low that they are not statistically reliable.

- Human Exposure to Contamination: contaminated sites in areas of human exposure concern. There are only three types of sites: Potential of interest, no interest of exposure, and insufficient data.

Report User:

- Regional data scientists will use the report as a starting point for a more comprehensive assessment of environmental trends.

- Upper management is not familiar with the details of the data set, but will use the report to monitors key environmental trends and will use created visualizations to communicate this information in external presentations.

- Policymakers who need to quickly understand data to support decisions about how to allocate capital to the aspects of the region that need their most attention.

7. Insurance Complaints Data

Scenario description: You are working at a large insurance company that is about to undergo an internal audit related to claims procedures. The company is forward thinking and now has a data warehouse along with several models. Management has asked you to review the model and provide details to management so that any issues can be highlighted and prepare them for the review. You must choose the analysis problems that you consider important.

Some factors to consider:

- The most recent 2 years of data are what stakeholders care about most

- Ability to view status changes and when they occur

- Complaints are broken down according to dimensions

- Customer satisfaction

- Brokers commit the most serious violations

Dataset description: 12 data tables, including 2 fact tables Complains Data and Status History Data, in addition there are dimension tables to describe and clarify the data.

8. HR Analytics Case Study

A company called XYZ, at any given time, employs about 4,000 people. However, every year, about 15% of a company's employees leave the company and need to be replaced by available talent in the job market. Management believes that this level of attrition (employees leaving voluntarily or due to layoffs) is bad for the company for the following reasons:

- Projects of former employees are behind schedule, difficult to meet schedules, causing loss of reputation with consumers and partners.

- Need to maintain a large department for the purpose of recruiting new talent

- Usually, new employees must be trained on the job and/or have time to get acquainted with the company.

Dataset description:

- Table employee_survey_data.csv: employee survey results include: NV code, Level of satisfaction with the working environment, Level of work engagement, Level of work-life balance.

- Table general_data.csv: general information about employees including: Age, Attrition, Business trip, Department, Distance from home to company, Education, Field of education, Employee code, Gender.

- Table in_time.csv: detailed time attendance to the employee's company including a number of working days in January 2015.
- Table manager_survey_data.csv: survey employee feedback about their manager including: NV Code, Work Participation, Performance Evaluation.
- Table out_time.csv: detailed time attendance of employees including a number of working days in January 2015.

By analyzing this dataset, companies will know what changes they should make in the workplace to keep the most employees there. Additionally, they want to know which factors have the biggest impact on employee churn and need to be addressed immediately.

9. Brazilian E-Commerce Public Dataset by Olist

Welcome! This is a Brazilian ecommerce public dataset of orders made at Olist Store. The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. We also released a geolocation dataset that relates Brazilian zip codes to lat/lng coordinates.

This is real commercial data, it has been anonymised, and references to the companies and partners in the review text have been replaced with the names of Game of Thrones great houses

10. Teco Customer churn

The data set describes the characteristics of customers of Telco company - a business providing telecommunications services and accompanying value-added services. In the company's customer base, a number of customers have left the service. Our goal is to help companies identify the factors that influence the likelihood of leaving their services. The company has strategies to retain customers.

Dataset description: The data set contains 7043 rows corresponding to 7043 customers, 21 columns.

The general goal is to find factors that affect customers' ability to leave. Some analysis directions are as follows:

- Portrait of Telco's customers
- Portrait of customers leaving Telco
- Criteria that influence the likelihood of leaving