CAPITAL NUMBERS™

End-to-End Data Lifecycle Management

DATA ENGINEERING

What We Do

DATA INGESTION

Extraction of structured, unstructured data coming from streaming and batch sources and refining/cleansing data to make it available on legacy database systems or cloud systems, to data scientists and business users for exploration and analysis

DATA STORAGE & ELT / ETL

Extracting, processing, transforming, and loading data techniques into various relational, non-relational, noSQL, big data systems and/or cloud storages, depending on data availability, volume, velocity, type of data

DATA MODERNIZATION

An efficient and smart approach for migrating business data to/from on-prem legacy systems into cloud storage infrastructure or new target platforms

What We Do

DATA PIPELINES

Building production-grade replayable and independent data workflow pipelines to move, transform and store data using various legacy, big data and / or cloud orchestration and data management pipeline tools and techniques like DF, Databricks, Synapse, Informatica, etc., to process data in batch and real time

DATA CI/CD

Expertise in legacy and cloud-based deployment services for developing efficient production build and release pipelines based on infrastructure-as-code artifacts, reference / application data, database objects (schema definitions, functions, stored procedures, etc.), data pipeline definitions and data validation and transformation logics

REAL-TIME PROCESSING

Expertise in implementing real-time and batch data processing systems across distributed environments based on mobile, web hosting and cloud services

Our Methodology

We use a consultative approach that combines data engineering, cloud, data privacy, and compliance expertise with proprietary frameworks and maturity models to construct a modern data ecosystem.

In addition, our flexible resourcing model allows for the rapid scaling of teams through a pod or virtual pod-based approach.

Challenges to Which We Provide Solutions

- Being able to turn a fast-growing pool of enterprise data into actionable intelligence

- A trustworthy data foundation and enabling analytics supported by insights from a wide range of data sources

- Proper data preparation that allows insights from raw data — for all types of analytics. Insights that are available in context-specific patterns for interactive visualizations, and predictive and prescriptive analytics

Some Data Engineering Technologies We Work With

Google Cloud

aws

Azure

databricks

Microsoft® SQL Server®

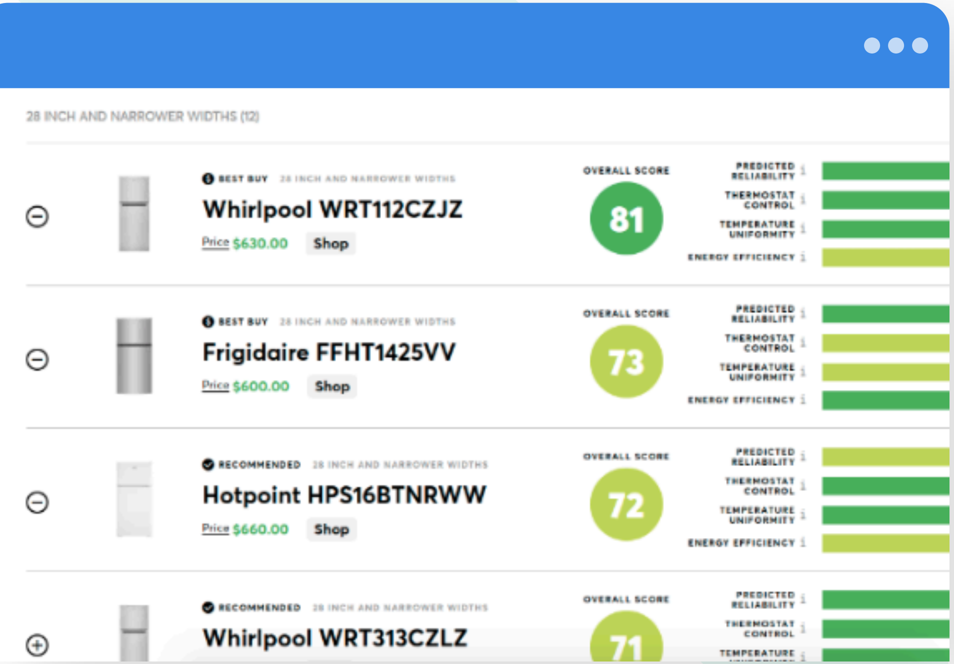
kafka

CASE STUDIES

Consumer Reports

Consumer Reports, is an American nonprofit consumer organization dedicated to independent product testing, investigative journalism, consumer-oriented research, public education, and consumer advocacy.

The motto is to provide trusted ratings and reviews for the products and services people use every day - from cars and major appliances to must-have tech gadgets, to home and garden necessities. Consumer Reports buys everything they test and doesn’t accept advertising. Users get the unbiased information they need to save time and money and have peace of mind. CR conducts considerable research — gathering data about products and services, consumer demand in the marketplace, and what their members care about most.



28 INCH AND NARROWER WIDTHS (12)

BEST BUY 28 INCH AND NARROWER WIDTHS
Whirlpool WRT112CZJZ
Price $630.00 Shop
OVERALL SCORE 81
PREDICTED RELIABILITY
THERMOSTAT CONTROL
TEMPERATURE UNIFORMITY
ENERGY EFFICIENCY

BEST BUY 28 INCH AND NARROWER WIDTHS
Frigidaire FFHT1425VV
Price $600.00 Shop
OVERALL SCORE 73
PREDICTED RELIABILITY
THERMOSTAT CONTROL
TEMPERATURE UNIFORMITY
ENERGY EFFICIENCY

RECOMMENDED 28 INCH AND NARROWER WIDTHS
Hotpoint HPS16BTNRWW
Price $660.00 Shop
OVERALL SCORE 72
PREDICTED RELIABILITY
THERMOSTAT CONTROL
TEMPERATURE UNIFORMITY
ENERGY EFFICIENCY

RECOMMENDED 28 INCH AND NARROWER WIDTHS
Whirlpool WRT313CZLZ
OVERALL SCORE 71
PREDICTED RELIABILITY
THERMOSTAT CONTROL
TEMPERATURE

Consumer Reports

Data on product reliability and owner satisfaction is gathered through regular surveys of Consumer Reports members. In these surveys, respondents answer questions about the products they own—in that way, Consumer Reports members are true partners in the product rating process! There are more than 6 million members who help in these surveys.

If you see the site https://www.consumerreports.org/ you will be able to see a lot of appliances for which detailed reviews and specifications are given. These data are coming from various sources and once we get that we run various converters which convert all the available files in the required format with required operations.

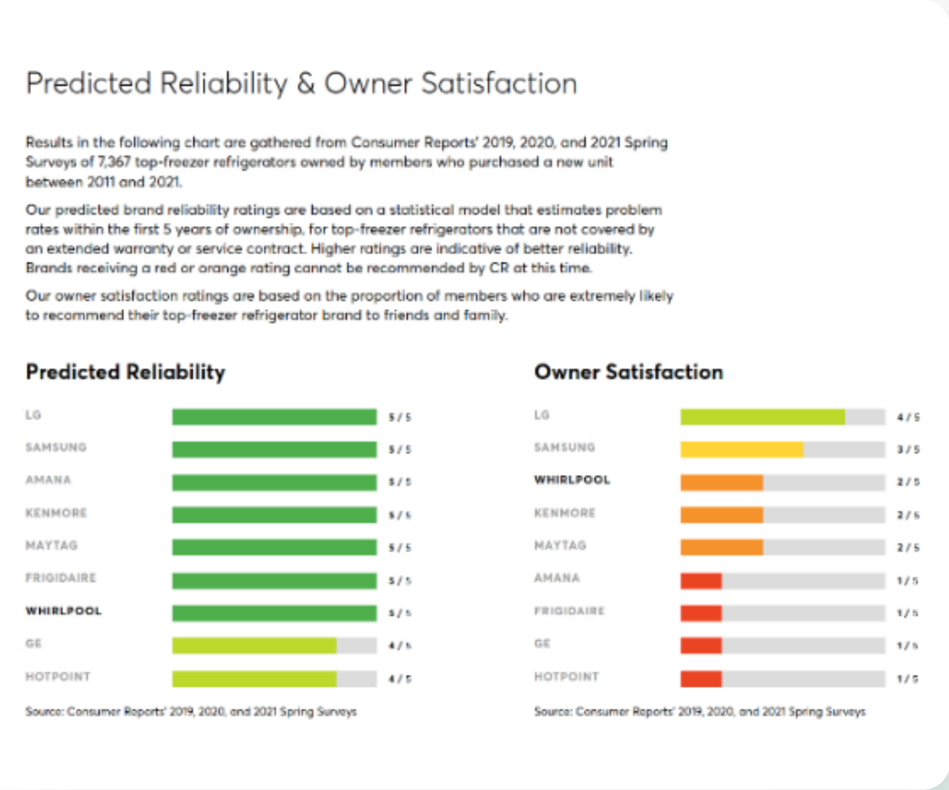
The data engineering team then creates different APIs, and the website consumes the information, and the data is shown beautifully in the frontend.

Consumer Reports

Sophisticated statistical models are used to calculate a brand's predictive reliability rating (reflecting estimated problem rates) in a given product category based on product age (measured in years purchased), frequency of use, and extended warranty or service contract Calculate as a function of range. We control for these factors in our model to ensure objective comparisons between brands.

These predictions are done based on the historical data from various past surveys in which the members participated.

In the following screenshot you can see the different Predicted Reliability scores for various refrigerator brands.



Predicted Reliability & Owner Satisfaction

Results in the following chart are gathered from Consumer Reports' 2019, 2020, and 2021 Spring Surveys of 7,367 top-freezer refrigerators owned by members who purchased a new unit between 2011 and 2021.

Our predicted brand reliability ratings are based on a statistical model that estimates problem rates within the first 5 years of ownership, for top-freezer refrigerators that are not covered by an extended warranty or service contract. Higher ratings are indicative of better reliability. Brands receiving a red or orange rating cannot be recommended by CR at this time.

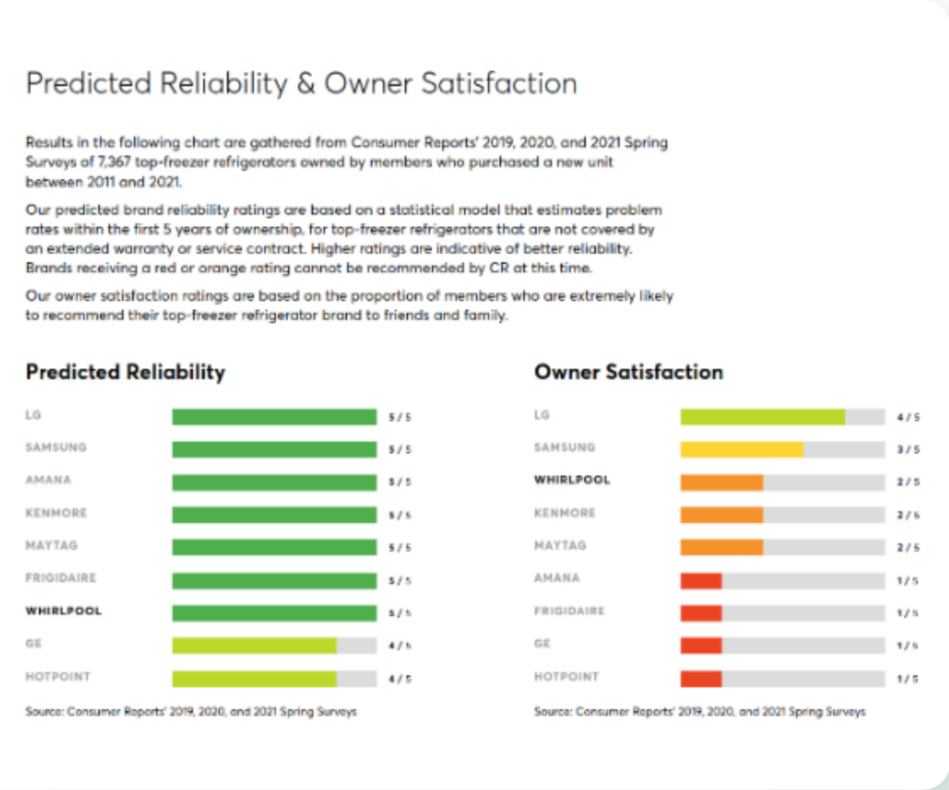
Our owner satisfaction ratings are based on the proportion of members who are extremely likely to recommend their top-freezer refrigerator brand to friends and family.

Predicted Reliability

Source: Consumer Reports' 2019, 2020, and 2021 Spring Surveys

Owner Satisfaction

Source: Consumer Reports' 2019, 2020, and 2021 Spring Surveys



Predicted Reliability & Owner Satisfaction

Results in the following chart are gathered from Consumer Reports' 2019, 2020, and 2021 Spring Surveys of 7,367 top-freezer refrigerators owned by members who purchased a new unit between 2011 and 2021.

Our predicted brand reliability ratings are based on a statistical model that estimates problem rates within the first 5 years of ownership, for top-freezer refrigerators that are not covered by an extended warranty or service contract. Higher ratings are indicative of better reliability. Brands receiving a red or orange rating cannot be recommended by CR at this time.

Our owner satisfaction ratings are based on the proportion of members who are extremely likely to recommend their top-freezer refrigerator brand to friends and family.

Predicted Reliability

LG	5 / 5
SAMSUNG	5 / 5
AMANA	5 / 5
KENMORE	5 / 5
MAYTAG	5 / 5
FRIGIDAIRE	5 / 5
WHIRLPOOL	5 / 5
GE	4 / 5
HOTPOINT	4 / 5

Source: Consumer Reports' 2019, 2020, and 2021 Spring Surveys

Owner Satisfaction

LG	4 / 5
SAMSUNG	3 / 5
WHIRLPOOL	2 / 5
KENMORE	2 / 5
MAYTAG	2 / 5
AMANA	1 / 5
FRIGIDAIRE	1 / 5
GE	1 / 5
HOTPOINT	1 / 5

Source: Consumer Reports' 2019, 2020, and 2021 Spring Surveys

Consumer Reports

Our Data engineering team also upgraded the Spring Boot version for CarsDMA. Following are the things we did:

- Spring Boot v1.4 to v2.3.12 Upgrade

- Major changes

  - Spring Boot Data Source Changes

  - JPA methods changes

  - MongoDB changes and upgrade (DB Object to Document)

  - Xml configuration to Java Bean convert

Buildings Alive

Buildings Alive, headquartered in Sydney, New South Wales, which helps owners and operators of large complex buildings to achieve their energy and environmental efficiency goals by cutting carbon emissions and cutting costs.

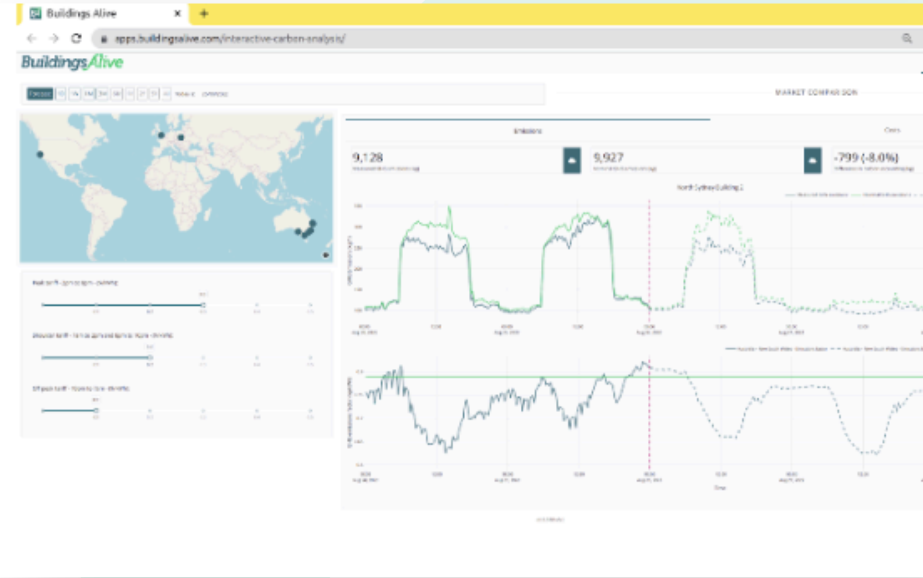
The main objective is to give owners, operators, engineers and technicians a clear understanding of how their buildings use energy and other resources by providing timely and unambiguous information with the help of data science

Buildings Alive

At present we have different ways to capture the consumption data of each building, which are as under:

- FTP
- Email
- API
- Direct files in CSV/Excel format

The data we collect is Energy, Water and Gas consumption data. This data is available for 15mins interval for everyday we will get 96 rows of consumption data for each building. Also, each building can have multiple meters providing multiple readings 96 rows of each meter.

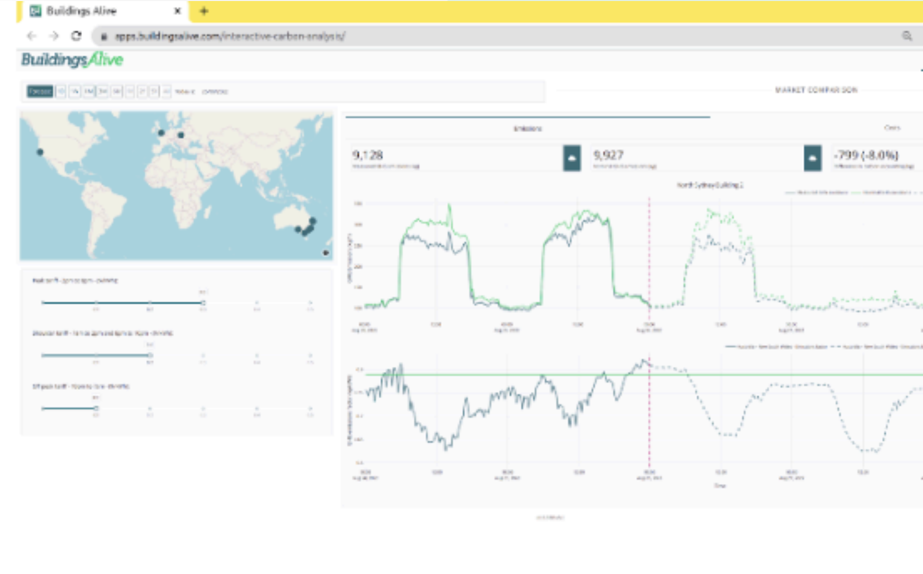


Buildings Alive

Once we got these files, we have written converters python scripts, which convert all the available files in the required format with required operations.

After reformatting and completing required operations, the data gets written into the database. The data is then associated with each building and each meter. This way we do the ETL (extract, transform and load) data integration processes.

Also based on the past data we are creating forecasts related to consumption and carbon emissions. We are showing these in a user-friendly dashboard.



Buildings Alive

Based on the processed data we have created 2 data marts. Data mart means joining multiple tables to create a single table which then can be directly used for specific purposes.
One is for state related data, and another is for building related data.

Using these data marts and some of the individual tables, we have created the dashboard where we show different charts like:

- Consumptions charts
- Carbon emission charts
- Forecast charts etc.

Based on these charts, users (owners and operators of large complex buildings) can see where consumption is higher, how they should manage consumption, and what they can expect in terms of consumption and carbon emissions in the coming days.

Every day the Data volume generated is around ~1,00,000 rows. This includes different meters data i.e., for water (35 million), gas (10 million) and energy (330 million). It includes carbon intensity data collected state-wise.

Buildings Alive

If we talk about Data variety, then the data is mainly of the numeric data type. In case any new building and meter is going to be added in that case it will contain textual data i.e., address, name, boolean etc.

For Data analysis and prediction models (AI/ML/Deep Learning), at present we are doing time series forecasting. Based on the past consumption i.e meter reading, we are predicting what the consumption will be for the next 5 days.

As we are working with the time series forecast, we are using line charts for the Data visualization. The charts are in real time.

We have various data sources, and the velocity of that data depends on provider to provider. The data range is from 5 minutes to 1 hour, which means that meter reading data will be available at every 5-minute interval.

The raw data that we collect needs to be converted to useful data using converters. It includes checking missing data for some interval, checking text data for random/unwanted text, etc. After the raw data goes through the converter, it is added to the database without any issue. This way, we take care of Data veracity as well.

About Capital Numbers

Capital Numbers is an award-winning Digital Consulting & Engineering Firm offering end-to-end software development solutions to Enterprises, Silicon Valley Founders, Digital Agencies, and Startups worldwide.

We are **ISO 9001** and **27001 certified** with **700+ experts** working full-time across multiple delivery centers and offices in India, US, and Australia.

With **259+ clients worldwide, 50+ awards, and 200+ five-star ratings,** Capital Numbers is currently ranked #1 on Clutch, G2, Trustpilot, and GoodFirms.

Thank You

We would love to hear from you

info@capitalnumbers.com

CN CAPITAL™ NUMBERS

capitalnumbers.com