Package ‘contoso’

March 1, 2026

Type Package

Title Dataset of the 'Contoso' Company

Version 2.1.0

Description A collection of synthetic datasets simulating sales transactions from a fictional company. The dataset includes various related tables that contain essential business and operational data, useful for analyzing sales performance and other business insights. Key tables included in the package are:
- ``sales'': Contains data on individual sales transactions, including order details, pricing, quantities, and customer information.
- ``customer'': Stores customer-specific details such as demographics, geographic location, occupation, and birthday.
- ``store'': Provides information about stores, including location, size, status, and operational dates.
- ``orders'': Contains details about customer orders, including order and delivery dates, store, and customer data.
- ``product'': Contains data on products, including attributes such as product name, category, price, cost, and weight.
- ``calendar'': A time-based table that includes date-related attributes like year, month, quarter, day, and working day indicators.
This dataset is ideal for practicing data analysis, performing time-series analysis, creating reports, or simulating business intelligence scenarios.

License MIT + file LICENSE

Imports DBI, dplyr, cli, duckdb (>= 1.4.0), purrr

Suggests testthat (>= 3.0.0), dbplyr, gt

Encoding UTF-8

LazyData true

RoxygenNote 7.3.3

Depends R (>= 4.1.0)

URL https://codeberg.org/usrbinr/contoso

Config/testthat/edition 3

BugReports https://codeberg.org/usrbinr/contoso/issues

NeedsCompilation no

Author Alejandro Hagan [aut, cre]

Maintainer Alejandro Hagan <alejandro.hagan@outlook.com>

Repository CRAN

Date/Publication 2026-03-01 03:30:08 UTC

Contents

`calendar` *Calendar Dimension Data from the Contoso Dataset*

Description

This dataset contains calendar-related information used for time-based analysis in the Contoso dataset. It includes various representations of date-related attributes, such as year, quarter, month, and day, along with indicators for working days. It is useful for time-series analysis and aggregating data by different time periods.

Usage

```
calendar
```

Format

A data frame with 17 columns:

date `Date` The actual date for the record.

date_key `double` Unique identifier for the date (often in YYYYMMDD format).

year `double` The year part of the date.

year_quarter `character` The year and quarter (e.g., "2025 Q1").

year_quarter_number `double` The numerical representation of the quarter (e.g., 1, 2, 3, 4).

quarter `character` The quarter of the year (e.g., "Q1", "Q2").

year_month `character` The year and month (e.g., "2025-03").

year_month_short `character` A shortened version of year and month (e.g., "2025 Mar").

year_month_number `double` The numerical representation of the year-month (e.g., 202503 for March 2025).

month `character` The month name (e.g., "March").

month_short `character` The abbreviated month name (e.g., "Mar").

month_number `double` The numerical representation of the month (e.g., 3 for March).

day_of_week `character` The full name of the day of the week (e.g., "Monday").

day_of_week_short `character` The abbreviated day of the week (e.g., "Mon").

day_of_week_number `double` The numerical representation of the day of the week (e.g., 1 for Monday).

working_day `double` Indicator of whether the date is a working day (1 for working day, 0 for non-working day).

working_day_number `double` A numerical indicator for working day (e.g., 1 for working day, 0 for non-working day).

Source

https://github.com/sql-bi/Contoso-Data-Generator-V2-Data/releases/tag/ready-to-use-data

---

`create_contoso_duckdb` *Creates DuckDB database with Contoso datasets*

Description

Creates a DuckDB connection with Contoso datasets loaded from cloud storage. The datasets are stored as Parquet files on Cloudflare R2 and streamed directly into DuckDB.

Usage

```
create_contoso_duckdb(size = "small")
```

Arguments

`size`	Dataset size: "small", "medium", "large", or "mega"

Details

The `create_contoso_duckdb()` function creates views for the following Contoso datasets:

- `sales`: Contains sales transaction data.
- `product`: Contains details about products, including attributes like product name, manufacturer, and category.
- `customer`: Contains customer demographic and geographic information.

- `store`: Contains information about store locations and attributes.
- `fx`: Contains foreign exchange rate data for currency conversion.
- `calendar`: Contains various date-related information, including day, week, month, and year.
- `orders`: Contains order header information.
- `orderrows`: Contains order line items.

Available sizes (approximate sales rows):

- `small`: ~8,000 rows
- `medium`: ~2.3 million rows
- `large`: ~47 million rows
- `mega`: ~237 million rows

Value

A list containing:

- `sales`, `product`, `customer`, `store`, `fx`, `calendar`, `orders`, `orderrows`: lazy `tbl` objects
- `con`: the DuckDB connection (use `DBI::dbDisconnect(db$con, shutdown = TRUE)` when done)

Examples

```
## Not run:
  db <- create_contoso_duckdb(size = "small")
  db$sales |> head()
  DBI::dbDisconnect(db$con, shutdown = TRUE)

## End(Not run)
```

---

customer *Customer Data from the Contoso Dataset*

Description

This dataset contains information about customers from the Contoso dataset, including demographic details, geographical information, contact information, and other personal attributes. It provides insights into customer profiles, including location, age, occupation, and more.

Usage

```
customer
```

Format

A data frame with 24 columns:

customer_key `double` Unique identifier for each customer.

geo_area_key `double` Unique identifier for the geographical area the customer resides in.

start_date `Date` Date when the customer relationship began.

end_date `Date` Date when the customer relationship ended, if applicable.

continent `character` The continent where the customer resides.

gender `character` The gender of the customer (e.g., 'Male', 'Female').

title `character` The title of the customer (e.g., 'Mr.', 'Ms.').

given_name `character` The given (first) name of the customer.

middle_initial `character` The middle initial of the customer, if applicable.

surname `character` The surname (last name) of the customer.

street_address `character` The street address of the customer.

city `character` The city where the customer resides.

state `character` The state or province where the customer resides.

state_full `character` The full name of the state or province.

zip_code `character` The postal (ZIP) code of the customer's address.

country `character` The country where the customer resides, using the country code.

country_full `character` The full name of the country where the customer resides.

birthday `Date` The date of birth of the customer.

age `double` The age of the customer.

occupation `character` The customer's occupation or profession.

company `character` The company the customer is associated with, if applicable.

vehicle `character` The type or make of vehicle the customer owns or drives.

latitude `double` The latitude of the customer's address.

longitude `double` The longitude of the customer's address.

Source

https://github.com/sql-bi/Contoso-Data-Generator-V2-Data/releases/tag/ready-to-use-data

fx *Foreign Exchange Data from the Contoso Dataset*

Description

This dataset contains information about foreign exchange (FX) rates between different currencies. It includes details about the exchange rate for a given date, as well as the currencies involved. This dataset is useful for analyzing currency conversions and understanding the exchange rates between different currencies over time.

Usage

```
fx
```

Format

A data frame with 4 columns:

date `Date` The date of the exchange rate.

from_currency `character` The code of the source currency (e.g., "USD", "EUR").

to_currency `character` The code of the target currency (e.g., "GBP", "JPY").

exchange `double` The exchange rate between the source and target currencies on the given date.

Source

https://github.com/sql-bi/Contoso-Data-Generator-V2-Data/releases/tag/ready-to-use-data

launch_ui *Launch the DuckDB UI in your browser*

Description

The `launch_ui()` function installs and launches the DuckDB UI extension for an active DuckDB database connection. This allows users to interact with the database via a web-based graphical interface.

Your connection from `create_contoso_duckdb()` is returned in the list.

Usage

```
launch_ui(.con)
```

Arguments

`.con`	A valid `DBIConnection` object connected to a DuckDB database. The function will check that the connection is valid before proceeding.

Details

The function performs the following steps:

- Checks that the provided DuckDB connection is valid. If the connection is invalid, it aborts with a descriptive error message.
- Installs the `ui` extension into the connected DuckDB instance.
- Calls the `start_ui()` procedure to launch the DuckDB UI in your browser.

This provides a convenient way to explore and manage DuckDB databases interactively without needing to leave the R environment.

Value

The function is called for its side effects and does not return a value. It launches the DuckDB UI and opens it in your default web browser.

See Also

- `create_contoso_duckdb()` for creating example Contoso datasets in DuckDB.
- `DBI::dbConnect()` and `DBI::dbDisconnect()` for managing DuckDB connections.
- `duckdb::duckdb()` for creating a DuckDB driver instance.

Examples

```
## Not run:
# Connect to DuckDB
db <- create_contoso_duckdb()

# Launch the DuckDB UI
launch_ui(db$con)

# Clean up
DBI::dbDisconnect(db$con, shutdown = TRUE)

## End(Not run)
```

---

orderrows *Order Rows Data from the Contoso Dataset*

Description

This dataset contains detailed information about the individual items (rows) within each order in the Contoso dataset. It includes details such as the product, quantity, pricing, and cost of each item in an order. This dataset is useful for analyzing the breakdown of order components and individual product sales.

Usage

```
orderrows
```

Format

A data frame with 7 columns:

order_key `double` Unique identifier for the order to which the item belongs.

line_number `double` Line number within the order, identifying each product line.

product_key `double` Unique identifier for the product in the order row.

quantity `double` The quantity of the product ordered.

unit_price `double` The price per unit of the product.

net_price `double` The total net price for the product, considering any applicable discounts.

unit_cost `double` The cost per unit of the product.

Source

https://github.com/sql-bi/Contoso-Data-Generator-V2-Data/releases/tag/ready-to-use-data

---

`orders` *Order Data from the Contoso Dataset*

Description

This dataset contains information about customer orders, including order dates, delivery dates, and store details.

Usage

```
orders
```

Format

A data frame with 6 columns:

order_key `double` Unique identifier for the order.

customer_key `double` Unique identifier for the customer who placed the order.

store_key `double` Unique identifier for the store where the order was placed.

order_date `Date` The date when the order was placed.

delivery_date `Date` The date when the order is expected to be delivered.

currency_code `character` The currency code used for the order (e.g., USD, EUR).

Source

https://github.com/sql-bi/Contoso-Data-Generator-V2-Data/releases/tag/ready-to-use-data

`product` *Product Data from the Contoso Dataset*

Description

This dataset contains information about products in the Contoso dataset. It includes product details such as identifiers, descriptions, pricing, weight, and categorization. This dataset is useful for analyzing product characteristics, pricing, and product-related sales insights.

Usage

```
product
```

Format

A data frame with 14 columns:

product_key `double` Unique identifier for each product.

product_code `character` A code that uniquely identifies the product.

product_name `character` The name or description of the product.

manufacturer `character` The name of the manufacturer of the product.

brand `character` The brand of the product.

color `character` The color of the product.

weight_unit `character` The unit of measurement for the product's weight (e.g., "kg", "lbs").

weight `double` The weight of the product.

cost `double` The cost price of the product.

price `double` The selling price of the product.

category_key `double` Unique identifier for the category to which the product belongs.

category_name `character` The name of the category to which the product belongs.

sub_category_key `double` Unique identifier for the subcategory to which the product belongs.

sub_category_name `character` The name of the subcategory to which the product belongs.

Source

https://github.com/sql-bi/Contoso-Data-Generator-V2-Data/releases/tag/ready-to-use-data

`sales` *Sales Data from the Contoso Dataset*

Description

This dataset contains information about sales orders, including order details, pricing, and customer data from the Contoso dataset. It provides insights into the transactions that have occurred, including order dates, delivery dates, customer and store information, as well as product details.

Usage

```
sales
```

Format

A data frame with sales columns:

order_key `double` Unique identifier for each order.

line_number `double` Line number within the order (for multi-line orders).

order_date `Date` Date when the order was placed.

delivery_date `Date` Date when the order was delivered.

customer_key `double` Unique identifier for the customer who placed the order.

store_key `double` Unique identifier for the store where the order was placed.

product_key `double` Unique identifier for the product in the order.

quantity `double` The quantity of the product ordered.

unit_price `double` The price per unit of the product.

net_price `double` The total net price for the product, considering any discounts.

unit_cost `double` The cost per unit of the product.

currency_code `character` The currency code used for the transaction (e.g., USD, EUR).

exchange_rate `double` The exchange rate applied to the currency, if applicable.

gross_revenue `double` A product's unit_price multiplied by quantity.

net_revenue `double` A product's net_price multiplied by quantity.

unit_discount `double` A product's unit_price minus net_price.

discounts `double` A product's unit_discount multiplied by quantity.

cogs `double` Cost of goods sold. A product's unit_cost multiplied by quantity.

gross_margin `double` A product's net_revenue minus cogs.

unit_margin `double` A product's gross_margin divided by quantity.

Source

https://github.com/sql-bi/Contoso-Data-Generator-V2-Data/releases/tag/ready-to-use-data

`store` *Store Data from the Contoso Dataset*

Description

This dataset contains information about stores within the Contoso dataset. It includes details about the store's geographic location, operational status, and physical characteristics such as size and opening/closing dates. It provides insights into the store network of the company.

Usage

```
store
```

Format

A data frame with 11 columns:

store_key `double` Unique identifier for each store.

store_code `double` A code that uniquely identifies the store.

geo_area_key `double` Unique identifier for the geographical area where the store is located.

country_code `character` The country code where the store is located (e.g., "US", "DE").

country_name `character` The full name of the country where the store is located.

state `character` The state or province where the store is located.

open_date `Date` The date when the store was opened.

close_date `Date` The date when the store was closed, if applicable.

description `character` A description of the store (e.g., "Flagship store", "Outlet store").

square_meters `double` The physical size of the store in square meters.

status `character` The operational status of the store (e.g., "Open", "Closed").

Source

https://github.com/sql-bi/Contoso-Data-Generator-V2-Data/releases/tag/ready-to-use-data

Index