

Package ‘nyctaxi’

October 26, 2017

Title Accessing New York City Taxi Data

Version 0.0.1

Date 2017-10-24

Description New York City's Taxi and Limousine Commission (TLC) Trip Data <http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml> is a collection of green and yellow taxi trip data. The data was collected and provided by technology providers under the Taxicab & Livery Passenger Enhancement Programs <<http://www.nyc.gov/html/tlc/html/industry/shl.shtml>>. This package helps users to download New York City Trip Data.

Depends R (>= 3.2.3), etl

Imports DBI, dplyr, rlang, stringr

License CC0

LazyData true

RoxygenNote 6.0.1

Suggests dbplyr, testthat, knitr, lubridate, leaflet, rmarkdown, RSQLite

VignetteBuilder knitr

NeedsCompilation no

Author Ben Baumer [aut],
Wencong Priscilla Li [aut, cre],
Trang Le [aut]

Maintainer Wencong Priscilla Li <liwencong1995@gmail.com>

Repository CRAN

Date/Publication 2017-10-26 08:23:57 UTC

R topics documented:

etl_extract.etl_nyctaxi	2
get_file_path	3
green_2016_01_sample	4

 etl_extract.etl_nyctaxi

Extract NYC Taxi Trip Data from data from NYC Taxi & Limousine Commission

Description

Extract NYC Taxi Trip Data from data from NYC Taxi & Limousine Commission

Usage

```
## S3 method for class 'etl_nyctaxi'
etl_extract(obj, years = as.numeric(format(Sys.Date(),
"%Y")), months = 1:12, types = "yellow", ...)

## S3 method for class 'etl_nyctaxi'
etl_load(obj, years = as.numeric(format(Sys.Date(),
"%Y")), months = 1:12, types = "yellow", ...)

## S3 method for class 'etl_nyctaxi'
etl_transform(obj, years = as.numeric(format(Sys.Date(),
"%Y")), months = 1:12, types = "yellow", ...)
```

Arguments

obj	an etl object
years	a numeric vector giving the years. The default is the most recent year.
months	a numeric vector giving the months. The default is January to December.
types	a character vector giving the type of taxi trip data of yellow
...	arguments passed to smart_download and/or green . The default is yellow.

Details

extract NYC Yellow taxi trip data from Jan 2009 and Green taxi trip data from Aug 2013 data from NYC Taxi & Limousine Commission

load NYC Yellow taxi trip data from load directory into a sql database, the default is a sqlite database and/or green. The default is yellow.

transform NYC Yellow taxi trip data from raw directory to load directory and/or green. The default is yellow.

See Also

[etl_extract](#)
[etl_load](#)
[etl_transform](#)

Examples

```
## Not run:
taxi <- etl("nyctaxi", dir = "~/Desktop/nyctaxi/")
taxi %>%
  etl_extract(years = 2016, months = 1:2, types = c("yellow", "green")) %>%
  etl_transform(years = 2016, months = 1:2, types = c("yellow", "green")) %>%
  etl_load(years = 2016, months = 1:2, types = c("yellow", "green"))

## End(Not run)
```

get_file_path	<i>Utility function that generates file paths</i>
---------------	---

Description

Take in arguments common to all three functions (e.g. years, months, types) as well as the stem of the URL and return the matched file paths.

Usage

```
get_file_path(years, months, types, path)
```

Arguments

years	a numeric vector giving the years. The default is the most recent year.
months	a numeric vector giving the months. The default is January to December.
types	a character vector giving the type of taxi trip data of yellow
path	path to desired file

Examples

```
get_file_path(2017, 1:6, "yellow", "~/")
```

green_2016_01_sample *NYC Taxi trip data from NYC's Taxi and Limousine Commission (TLC)*

Description

Random samples of 1000 observations generated by the `sample` function from the 2016 January green and yellow taxi trip data.

Usage

```
green_2016_01_sample
```

```
yellow_2016_01_sample
```

Format

For both, a dataframe (`tbl_df`) each with 1,000 rows and 21 or 19 variables.

The following variables are values that are found in either the `green_2016_01_sample` or `yellow_2016_01_sample` data tables.

VendorID A code indicating the LPEP provider that provided the record. 1= Creative Mobile Technologies, LLC; 2= VeriFone Inc.

lpep_pickup_datetime The date and time when the meter was engaged.

lpep_dropoff_datetime The date and time when the meter was disengaged.

lpep_pickup_datetime Taxi pickup date and time

lpep_dropoff_datetime Taxi dropoff date and time

Store_and_fwd_flag This flag indicates whether the trip record was held in vehicle memory before sending to the vendor, aka "store and forward," because the vehicle did not have a connection to the server. Y= store and forward trip, N= not a store and forward trip

RareCodeID The final rate code in effect at the end of the trip. 1= Standard rate, 2=JFK, 3=Newark, 4=Nassau or Westchester, 5=Negotiated fare, 6=Group ride

Pickup_longitude Longitude where the meter was engaged.

Pickup_latitude Latitude where the meter was engaged.

Dropoff_longitude Longitude where the meter was timed off.

Dropoff_latitude Latitude where the meter was timed off.

Passenger_count The number of passengers in the vehicle. This is a driver-entered value.

Trip_distance The elapsed trip distance in miles reported by the taximeter.

Fare_amount The time-and-distance fare calculated by the meter.

Extra Miscellaneous extras and surcharges. Currently, this only includes the \$0.50 and \$1 rush hour and overnight charges.

MTA_tax \$0.50 MTA tax that is automatically triggered based on the metered rate in use.

Tip_amount Tip amount - This field is automatically populated for credit card tips. Cash tips are not included.

Tolls_amount Total amount of all tolls paid in trip.

Ehail_fee The description of this variable is not available on TLC website.

improvement_surcharge \$0.30 improvement surcharge assessed on hailed trips at the flag drop. The improvement surcharge began being levied in 2015.

Total amount The total amount charged to passengers. Does not include cash tips.

Payment_type A numeric code signifying how the passenger paid for the trip. 1= Credit card, 2= Cash, 3= No charge, 4= Dispute, 5= Unknown, 6= Voided trip

Trip_type A code indicating whether the trip was a street-hail or a dispatch that is automatically assigned based on the metered rate in use but can be altered by the driver. 1= Street-hail, 2= Dispatch

Details

NYC's Taxi and Limousine Commission (TLC) Trip Data is a collection of trip records including fields capturing pick-up and drop-off locations, times, trip distances, fares, rate types, and driver-reported passenger counts. The data was collected and provided to the NYC TLC by technology providers under the Taxicab & Livery Passenger Enhancement Programs.

Author(s)

Wencong (Priscilla) Li

Source

These data are recorded at http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

References

The description of the data is available on http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

Examples

```
## Not run:
#' #
if (require(dplyr) & require(ggplot2)) {
  green_2016_01_sample %>%
  filter(Trip_distance <= 1) %>%
  ggplot(aes(x = Dropoff_longitude, y = Pickup_latitude))
}

## End(Not run)
```

Index

*Topic **datasets**

- green_2016_01_sample, 4

- etl_extract, 2
- etl_extract.etl_nyctaxi, 2
- etl_load, 2
- etl_load.etl_nyctaxi
 - (etl_extract.etl_nyctaxi), 2
- etl_transform, 2
- etl_transform.etl_nyctaxi
 - (etl_extract.etl_nyctaxi), 2

- get_file_path, 3
- green_2016_01_sample, 4

- sample, 4
- smart_download, 2

- tbl_df, 4

- yellow_2016_01_sample
 - (green_2016_01_sample), 4