ISA 401/501: Business Intelligence & Data Visualization

06: Connecting to APIs in R

Fadel M. Megahed, PhD

Enders Associate Professor
Farmer School of Business
Miami University

🐦 @FadelMegahed
 fmegahed
✈ fmegahed@miamioh.edu
❓ Automated Scheduler for Office Hours

Fall 2022

Quick Refresher from Last Week [1]

✅ Understand when can we scrape data (i.e., `robots.txt`)

✅ Scrape a webpage Using ®R

✅ Utilize loops to scrape data from multiple webpages

Quick Refresher from Last Week [2]

A Note on HTML Element vs Elements for Tables

In office hours, I was asked a question related to the impact of using `[[]]` on the output from `html_table()` when `step2` utilizes `html_element()` vs `html_elements()`. So let us examine the Miami Football Wikipedia Page to: (a) answer this question; and (b) see the impact of `html_element()` vs `html_elements()` when there are multiple tables of the same class.

One more Example on Scraping Multiple Webpages

The most popular listings on Netflix are are available at https://www.imdb.com/search/title/?companies=co0144901. Write an ℝ script that will produce a tibble that contains the **following information for the first 300 entries**: (a)title, which you will save in a column titled `title`; (b)

Learning Objectives for Today's Class

- Describe what we mean by an API

- Explain how APIs will be a huge part of your career as a business analyst and/or data scientist

- Use APIs for extracting web data

What is an API?
(A Web Server Bases Perspective)

What is an API? [1]

- An **API** is an acronym for application programming interface.

- It is a **popular** approach to interact with an application/service or data since it:

  - Defines a set of functionalities independent of implementation (i.e., it only exposes information that a programmer might find useful and keep those parts consistent even if the implementation changes later)

  - Provides some level of privacy/control over one's internal data and the rate at which it can be accessed.

What is an API? [2]

What is an API? [3]

Scenario: Alone, you went into a warehouse and are trying to retrieve 3 screwdrivers, a toolbox, and 15 phillips screws. But you do not know, where those things are in the warehouse.

The API is the set of instructions provided by the warehouse manager on where/how to retrieve this information without touching/accessing other things in there.

Source: Matt Z. (2018). What is an API? (explanation with cartoon picture so a 5 year old could understand it)

API Useages for Business Analysts and Data Scientists

Application #1: Request Data from Web Server

This is the **classical** example, which was explained in the previous slides and we will be coding several in-class demos to illustrate how to perform such operations in Ⓡ.

Application #2: Access Python Libraries

Search the docs ...

Input/output ∧

    pandas.read_pickle

    pandas.DataFrame.to_pickle

    pandas.read_table

    pandas.read_csv

    pandas.DataFrame.to_csv

    pandas.read_fwf

    pandas.read_clipboard

    pandas.DataFrame.to_clipboard

pandas.read_csv

pandas.read_csv(*filepath_or_buffer, sep=NoDefault.no_default, delimiter=None, header='infer', names=NoDefault.no_default, index_col=None, usecols=None, squeeze=None, prefix=NoDefault.no_default, mangle_dupe_cols=True, dtype=None, engine=None, converters=None, true_values=None, false_values=None, skipinitialspace=False, skiprows=None, skipfooter=0, nrows=None, na_values=None, keep_default_na=True, na_filter=True, verbose=False, skip_blank_lines=True, parse_dates=None, infer_datetime_format=False, keep_date_col=False, date_parser=None, dayfirst=False, cache_dates=True, iterator=False, chunksize=None, compression='infer', thousands=None, decimal='.', lineterminator=None, quotechar='"', quoting=0, doublequote=True, escapechar=None, comment=None, encoding=None, encoding_errors='strict', dialect=None,*

Application #3: Use a Pretrained ML Model

Sorry

Because of its privacy settings, this video cannot be played here.
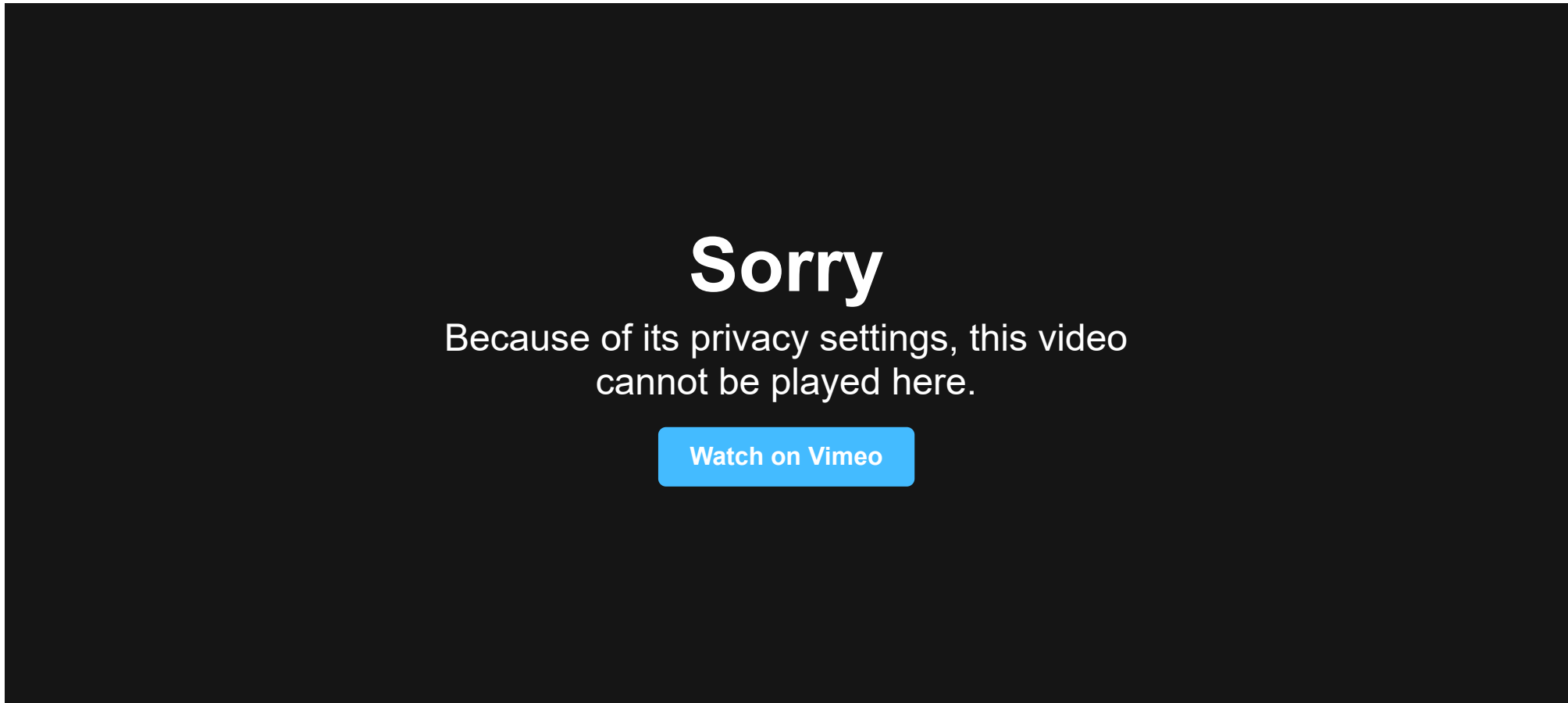
Watch on Vimeo

Please click on the "Watch on Vimeo" to see OpenAI's explanation of their DALL·E 2 model.

Application #4: Deploy a ML Model

Predicting In-Hospital Mortality After TAVR Using Preoperative Variables and Penalized Logistic Regression

The app below utilizes the finalized logistic regression model with an l2 penalty based on the manuscript by Alhwiti et al. The manuscript will be submitted to JACC: Cardiovascular Interventions. The data used for model building is all TAVR procedures between 2012 and 2019 as reported in the HCUP NIS database.

The purpose of the app is to provide evidence-based clinical support for interventional cardiology. .

Age	60

output

Sex

Source: To access the app outside of the slides, please visit fmegahed/tavr_project.

Accessing APIs in R

The 3 Step Process

Before you dive into the API documentation, you **should first check if there is a R (or Python if you are familiar)** package/library that serves as a wrapper for that API.

- **Find** the **API's documentation** and find information about the following:
  A. Does the API require an **authentication key**?
  B. What are the API's **base URL** and **query parameters**?
  C. How does the request URL look like?

- Craft your **request**. My recommendations are to:
  A. First, start with a simple request.
  B. Test that request in your browser and see what results you get.

- Construct that request in ® by **either**:
  A. If the generated content seems to be a `JSON` file/webpage, you can capitalize on the reading the content from `jsonlite::fromJSON()`; **OR**

Demo 1: `tidycensus` vs Census API

Motivation   Learning Objectives   tidycensus Results   Direct Results

In socio-economic analysis, we are often interested in examining explanatory population-level variables. For the U.S., the decennial (once every 10 ten years) Census, and the 1-year and 5-year American Community Surveys are often the gold-standard for such data.

Luckily, the tidycensus is an R package that allows users to interface with a select number of the US Census Bureau's data APIs and return tidyverse-ready data frames.

Demo 1: `tidycensus` vs Census API

In this demo, we will:

- Set up an API key for the Census API

- Use the `tidycensus` package to obtain the total population for Butler and Warren Counties in Ohio from https://api.census.gov/data/2020/dec/pl/variables.html

- Extract the same data by capitalizing on the API itself (i.e., without the tidycensus package)

Demo 1: `tidycensus` vs Census API

In class, we will live code and capitalize on the `tidycensus` package to get the total population for Butler and Warren Counties in Ohio from
https://api.census.gov/data/2020/dec/pl/variables.html

```
## # A tibble: 2 × 4
##   GEOID NAME                 variable  value
##   <chr> <chr>                <chr>     <dbl>
## 1 39017 Butler County, Ohio P1_001N  390357
## 2 39165 Warren County, Ohio P1_001N  242337
```

Demo 1: `tidycensus` vs Census API

Motivation   Learning Objectives   tidycensus Results   **Direct Results**

> In class, we will live code and capitalize on the `tidycensus` package to get the total population for Butler and Warren Counties in Ohio from
> https://api.census.gov/data/2020/dec/pl/variables.html

```
## # A tibble: 2 × 3
##   P1_001N state county
##   <chr>   <chr> <chr>
## 1 390357  39    017
## 2 242337  39    165
```

Demo 2: Accuweather API

Demo Description Code and Results

- Go to https://developer.accuweather.com/ and create an account.
- Add your first app (from the `MY APPS` tab) and copy the generated API key.
- Then using the API Reference Tab ➜ Locations API ➜ City Search ➜ find the location key for `Oxford, Ohio`
- Use this information in the `Forecast API` to obtain the `5 Day Forecasts` for `Oxford Ohio`.

Demo 2: Accuweather API

Demo Description   **Code and Results**

```
## List of 2
##  $ Headline      :List of 9
##   ..$ EffectiveDate    : chr "2022-09-11T08:00:00-04:00"
##   ..$ EffectiveEpochDate: int 1662897600
##   ..$ Severity         : int 3
##   ..$ Text             : chr "Thunderstorms in the area this morning through this evenin
##   ..$ Category         : chr "thunderstorm"
##   ..$ EndDate          : chr "2022-09-12T02:00:00-04:00"
##   ..$ EndEpochDate     : int 1662962400
##   ..$ MobileLink       : chr "http://www.accuweather.com/en/us/oxford-oh/45056/daily-wea
##   ..$ Link             : chr "http://www.accuweather.com/en/us/oxford-oh/45056/daily-wea
##  $ DailyForecasts:'data.frame':    5 obs. of  8 variables:
##   ..$ Date         : chr [1:5] "2022-09-11T07:00:00-04:00" "2022-09-12T07:00:00-04:00" "202
##   ..$ EpochDate    : int [1:5] 1662894000 1662980400 1663066800 1663153200 1663239600
##   ..$ Temperature:'data.frame':    5 obs. of  2 variables:
##   .. ..$ Minimum:'data.frame':    5 obs. of  3 variables:
##   .. ..$ Maximum:'data.frame':    5 obs. of  3 variables:
##   ..$ Day          :'data.frame':    5 obs. of  5 variables:
##   .. ..$ Icon             : int [1:5] 15 6 4 1 2
```

Demo 3: The CryptoCompare API

Demo Description   Code and Results

- Create a Personal (Free) account at CryptoCompare.com

- Click on create your free key to create your API key and copy the key.

- Go to the documentation, and test their sample call by executing the call after you have pasted your API key in the call.

  - The executed call returns the price of BTC (Bitcoin) in USD, JPY and EUR.

- Now click on the Historical Data Tab on the left

  - Click on Daily Pair OHLCV and Execute the Sample Call for BTC

  - This returns 10 days worth of OHLCV for BTC in USD.

- Let us obtain the price for $SHIB over the past 100 days.

Demo 3: The CryptoCompare API

Demo Description **Code and Results**

```
##             time       high        low       open      close
## 96   2022-09-06 1.292e-05 1.172e-05 1.264e-05 1.202e-05
## 97   2022-09-07 1.239e-05 1.175e-05 1.202e-05 1.224e-05
## 98   2022-09-08 1.249e-05 1.198e-05 1.224e-05 1.222e-05
## 99   2022-09-09 1.394e-05 1.219e-05 1.222e-05 1.292e-05
## 100  2022-09-10 1.357e-05 1.272e-05 1.292e-05 1.327e-05
## 101  2022-09-11 1.350e-05 1.255e-05 1.327e-05 1.308e-05
```

Recap

Summary of Main Points

By now, you should be able to do the following:

- Describe what we mean by an API

- Explain how APIs will be a huge part of your career as a business analyst nad/or data scientist

- Use APIs for extracting web data

Things to Do to Prepare for Next Class

- Go over your notes, read through the supplementary material (below), go through the self-paced tutorial, and complete assignments 05 & 06 on Canvas.

- Getting Started with httr

- Managing secrets

- Per our tentative schedule, we will have an in-class lab where you will be expected to combine your new web scraping skills with APIs.