

ISA 401: Business Intelligence & Data Visualization

09: Connecting to APIs in

Fadel M. Megahed, PhD

Enders Associate Professor
Farmer School of Business
Miami University

 @FadelMegahed


 fmegahed

 fmegahed@miamioh.edu

 Automated Scheduler for Office Hours

Fall 2023

Quick Refresher from Last Week [1]

- ✓ Understand when can we scrape data (i.e., `robots.txt`)
- ✓ Scrape a webpage Using 
- ✓ 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.

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 Based 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.

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

 Getting started

User Guide

API reference

Development

Release notes

 2.1.1 

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

[Home](#) > [API reference](#) > [Input/output](#) > [pandas.read_csv](#)

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, 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=_NoDefault.no_default, keep_date_col=False, date_parser=_NoDefault.no_default, date_format=None, dayfirst=False,`

Application #3: Use a Pretrained ML Model



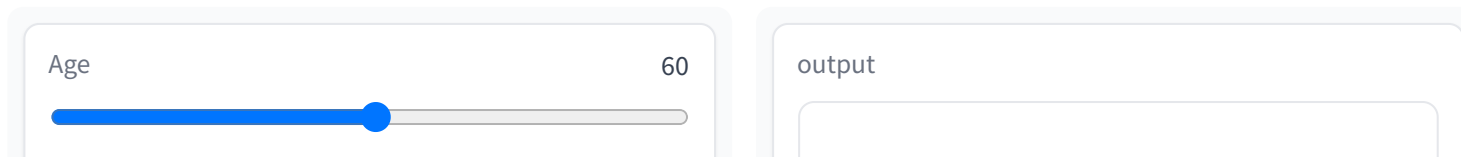
Please click on the ["Watch on Vimeo"](#) to see OpenAI's explanation of their DALL·E 2 model.

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, Aldrugh, and Megahed**. The manuscript is under review at Scientific Reports. The data used for model building is all TAVR procedures between 2012 and 2019, as reported in the HCUP NIS database.

The app's purpose is to provide evidence-based clinical support for interventional cardiology.



The screenshot shows a web application interface. On the left, there is a slider control labeled 'Age' with a blue handle positioned at the value '60'. To the right of the slider is a text input field labeled 'output'.

Accessing APIs in

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 **R** 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**
 - B. By passing the **base url** inside the `httr::GET()` and parsing the results with `httr::content()`.

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

Motivation

Learning Objectives

tidycensus Results

Direct Results

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

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 × 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
	<pre>## List of 2 ## \$ Headline :List of 9 ## ..\$ EffectiveDate : chr "2023-09-27T02:00:00-04:00" ## ..\$ EffectiveEpochDate: int 1695794400 ## ..\$ Severity : int 2 ## ..\$ Text : chr "Showers and thunderstorms around late tomorrow night throu ## ..\$ Category : chr "thunderstorm" ## ..\$ EndDate : chr "2023-09-28T20:00:00-04:00" ## ..\$ EndEpochDate : int 1695945600 ## ..\$ 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] "2023-09-25T07:00:00-04:00" "2023-09-26T07:00:00-04:00" "202 ## ..\$ EpochDate : int [1:5] 1695639600 1695726000 1695812400 1695898800 1695985200 ## ..\$ 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] 6 12 17 13 3 ## \$ IconPhrase : chr [1:5] "Mostly cloudy" "Showers" "Partly sunny w/ t-s</pre>

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		2023-09-20	27387.52	26821.46	27218.95	27126.17
## 97		2023-09-21	27159.51	26372.82	27126.17	26567.49
## 98		2023-09-22	26737.30	26481.86	26567.49	26583.27
## 99		2023-09-23	26640.71	26522.44	26583.27	26585.41
## 100		2023-09-24	26734.10	26149.51	26585.41	26262.42
## 101		2023-09-25	26421.98	26004.18	26262.42	26332.79

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), and complete [assignment 07](#) on Canvas.
- [Getting Started with http](#)
- [Managing secrets](#)