

Chapter 09 - Using APIs for Data Analytics

Custom Metrics for Analytics

Custom Metric = A calculation that summarizes complicated behavior, ability, and outcomes as a number

What is required?

- Question:
 - What question is it trying to solve? It should be general enough to have broad application but specific enough to add new knowledge to the field.
- Theory:
 - By choosing specific numbers to measure and weighing them against each other, you make a value judgment. You are proposing subcomponents that matter to answer a question.
- Valid Approach:
 - Do the underlying calculations support the purpose of the metric?
- Data Source:
 - Can you get data to calculate the metric at a reasonable frequency?
 - If data isn't available, your approach and the supporting theory may have to be adjusted out of practicality
- Name
 - The more interesting name = more impact can have

Using APIs as Data Sources for Fantasy Custom Metrics

- To have a data that is updated frequently for your own chart = automated process which has 2 ways:
 - APIs
 - Web Scrapping
 - Involves using program code to read the HTML from a website page and extract the data.
 - Problem: Every time website changes = web scraper breaks
- [Super Bowl - Custom Metric Competition](#)

Tools

- Backoff: Python library for adding backoff and retry to web calls
- HTTPX: Python library for making web calls
 - Similar to requests library but support asynchronous API (You can do other job even if the other one is not finished yet)
- Jupyter Notebook: Interactive data science environment
 - Enable Interactive Computing = Code Cell + Markdown Comment Cell + Results
- Pandas: Data Analysis and formatting library
 - Provide Python with new data type $\xrightarrow{\text{which is}}$ Data Frame \rightarrow Two-Dimensional Structure: Rows + Columns
 - Provide methods:
 - Data manipulation
 - Filtering
 - Formatting
 - Good Learning Resource for Panda = [Pandas Documentation](#)

Creating an API Client File

`swc_simple_client.py`: Standalone Python file to make all the API calls which you will be using in other Jupyter Notebooks

- Use `backoff` to implement "Exponential Backoff and retry with jitter" = make API call reliable
- HTTPX in a context manager style
- Standard Deviation
- Coefficient of variation = Standard Deviation / Mean $\xrightarrow{\text{why it works?}}$ it is dimensionless
 $\xrightarrow{\text{means}}$ can be compared across values of difference sizes
 - CV is lower if it varies less $\xrightarrow{\text{but}}$ We need metrics where a higher number is better
 $\xrightarrow{\text{solution}}$ multiply by 100 and subtract by 100

Additional Resource:

- [NBA Insider: Is It Numbers or Talent? Sorting Fact, Fiction in NBA Stats Wave](#)
- [Basic Syntax | Markdown Guide](#)
- [Learn to Code with Fantasy Football - Python for Fantasy Football Data](#)
- [Coefficient of Variation: Definition and How to Use It](#)

- [Python for Data Analysis, 2nd Edition \[Book\]](#)