**Project Title**

**McGill Football Analytics Automation Project**

**Table of Contents**

1. Description
2. Tech Assets and URLs
3. Installation
4. Usage
5. Features
6. Data Sources
7. Troubleshooting
8. Costs Involved
9. Passwords and Credentials
10. Contact Information

**1. Description**

This project was developed as part of a community collaboration with the McGill Football team. Our goal was to support the coaching staff—specifically Maxime and Chris—by building practical, automated solutions for both internal game analytics (Maxime’s stream) and external data integration (Chris’s stream).

The first stream (Maxime’s) focused on improving the efficiency and strategic value of in-game analysis using the DV Sport platform. This involved designing and delivering 10 automated reports based on game-tracked datasets. These reports helped Maxime extract insights related to zone targeting, coverage breakdowns, run efficiency, quarterback decision-making, and more—all formatted for easy reuse each week.

The second stream (Chris’s) aimed to set up a long-term infrastructure for integrating external data such as social media sentiment and third-party stats via APIs. A working pipeline was created using snscrape to collect Twitter data and a prototype pipeline was built for future full-API use. These were designed with modularity in mind to support future goals like scouting, recruiting, and performance benchmarking.

**2. Tech Assets and URLs**

* Excel Pivot Reports (10): Custom game analysis dashboards built from the DV Sport dataset.
* Heatmaps and Visual Dashboards: Generated for key zones, coverage schemes, and play outcomes.
* Python Scripts:
  + snscrape\_pipeline.py (Implemented)
  + twitter\_api\_pipeline.py (Prototype)
  + Data cleaning, configuration, and CSV export scripts
* Config File: Editable file for setting keywords, date ranges, and export paths
* Sample Dataset: DV Sport-exported Excel files tracking every play
* GitHub Repository (internal team only): [Link to be added by team]

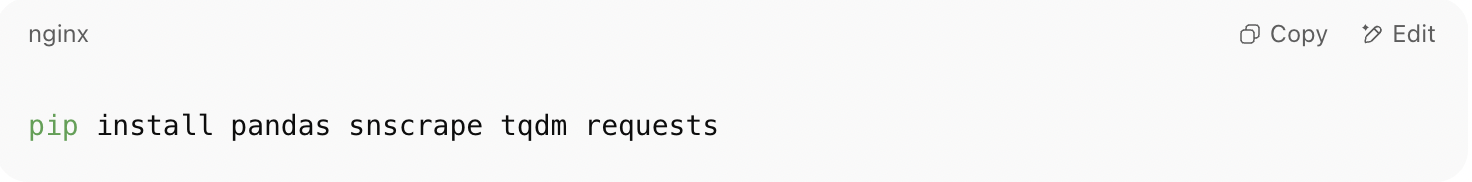
**3. Installation**

**For Maxime’s Excel Reports:**

No installation needed. Open Excel files and refresh pivot tables to update with new data.

**For Chris’s Twitter Pipeline:**

1. Clone or download the project folder.
2. Install Python 3.7+ and the following libraries:



1. Open the snscrape\_pipeline.py script and update the config variables.
2. Run the script to generate structured .csv files in the output folder.

**4. Usage**

**Maxime’s Reports (10 reports automated):**

1. Import new DV Sport Excel dataset to the workbook.
2. Refresh all pivot tables (right-click > Refresh).
3. Use slicers to filter by play type, QB, receiver, or coverage.
4. Visual outputs such as heatmaps and summaries are automatically updated.

**Chris’s Twitter Pipeline:**

1. Adjust the configuration for hashtags, accounts, and date ranges.
2. Run snscrape\_pipeline.py in a Python environment.
3. Review exported .csv containing structured tweet data.
4. Data can be loaded into Excel, Tableau, or Python-based models.

**5. Features**

**Maxime’s DV Sport Reports:**

* 10 structured pivot-based analytical reports
* Completion % by zone, coverage, receiver, and QB
* Run vs pass efficiency broken down by hash/formation
* Play gain buckets and gap-based run success
* Visual heatmaps for targeting and effectiveness
* Designed for non-technical reuse by coaches

**Chris’s External Integration Stream:**

* Real-time tweet scraping using snscrape
* Configurable keywords, hashtags, and account tracking
* Outputs structured .csvs ready for Excel/BI tools
* Prototype for Twitter API v2 integration
* Ready for expansion to CFL and Genius Sports APIs

**6. Data Sources**

* **Internal Data (DV Sport)**
  + Game-by-game Excel sheets tracking play number, yardage, down, formation, motion, coverage, result, QB, receiver, etc.
* **External Data (Chris’s Stream)**
  + Twitter (via snscrape)
  + Planned: CFL stats platform, Genius Sports APIs

**7. Troubleshooting**

**Maxime’s Reports:**

* If pivot tables show old data: Refresh all (Data tab > Refresh All)
* If slicers aren’t filtering: Ensure data range matches new dataset
* Formatting broken? Reimport original template file

**Chris’s Scripts:**

* Error: "snscrape not found" → Ensure snscrape is installed via pip
* Output not generated → Check date format in config
* Incomplete metadata → Use Twitter API version when available

**8. Costs Involved**

* All solutions built using free, open-source tools (Excel, Python, snscrape)
* Future costs may apply for full API access (e.g., Twitter Academic Access, Genius Sports credentials)
* Cloud hosting (optional for future automation) would incur minimal operational costs

**9. Passwords and Credentials**

* No credentials needed for the current implementation
* Twitter API v2 prototype will require a Bearer Token (not included)
* Future integrations with CFL/Genius will require secure credential handling

**10. Contact Information**

**Shuxi Chen**  
McGill University, Master of Management in Analytics  
📧 shuxi.chen@mail.mcgill.ca

**Hamza Javed**  
McGill University, Master of Management in Analytics  
📧 hamza.javed@mail.mcgill.ca

**Junpeng Kou**  
McGill University, Master of Management in Analytics  
📧 junpeng.kou@mail.mcgill.ca