1. **Introduction （400-500）**

Affected by the covid-19 , the public's demand for home entertainment has increased significantly. Playing games is also becoming indispensable for most young people under such social pressure. The happiest time for most modern young people is undoubtedly playing switch games on the sofa after work or study. This social situation has boosted Nintendo's game consoles and game software sales.

Nintendo is a Japanese company that mainly focuses on developing video game software and hardware. This company is one of the three majorities of the video game industry, and currently, it is the modern video game industry pioneer. Their most popular product is Switch, a console released in March 2017, which is a revolutionary console.

Switch's features are its design, which adopts the integrated design of the home console and handheld console. This design arguably redefines gaming. Players can play Switch games anywhere, such as at bus stops, planes, trains, or even in a doctor's waiting room due to its portability.

Five years past from 2017, the Switch has sold about 107.2 billion units in its lifetime. This impressive selling result made Nintendo Switch turn to be the seventh video game platform to sell more than 100 million units. Other platforms such as PS2, DS, Game Boy, PS4, PS1, and Wii are well-known. This indicates that Nintendo Switch is promoted to be one of the most popular products in the gaming world. A lot has changed for Nintendo since the Switch arrived in 2017; Nintendo Switch's derivatives have also kept up with trends and come into the public eye. One of the most important is Switch Games. Players can experience a variety of games with Switch, top switch games are simply stunning and profitable.

Metacritic is a website that collects reviews on games, which will integrate the scores of each reviews and make a final mark on the games. The final score is worthy as a reference by many players.

In this project, we will scrape data from Metacritic about the top hundred switch games and data from youtube. We aim to use those data to create a database, which will be convenient for further data pipeline study. >>>>我们要做什么 步骤 最后添加

Graph : Work flow create ([https://lucid.app/lucidchart/9b96a3e0-bd57-4acd-92b4-6d1e67ed96b7/edit?beaconFlowId=596B37B87BD2A753&page=0\_0&invitationId=inv\_ae3ec4e5-f8ca-41db-b267-12acbb4e7915#](https://lucid.app/lucidchart/9b96a3e0-bd57-4acd-92b4-6d1e67ed96b7/edit?beaconFlowId=596B37B87BD2A753&page=0_0&invitationId=inv_ae3ec4e5-f8ca-41db-b267-12acbb4e7915))

1. **Data Extraction**

I In this project, we used two methods to grab data to ensure the information is realistic and up to date in the database. The first method is called web scraping, which we used as the first web crawler and second web crawler. The web crawler is a simulation way as a browser to open a webpage to extract the data we want from that page. As long as the data can be accessed through a browser, we can acquire it through the web crawler. The first web we call here stands for the first web page we reach. The second web crawler stands the page click through the first page. API is the abbreviation of the Application Programming Interface as the second method we have used during this project to help in data extraction. An API is simply a programming code that will allow data transfer between software. API working processing is as shown in the following flow graph. This method is formal and under license, as most platforms provide API to the public. Both ways extract data with their unique benefits. We will show how we played with these two methods to catch data in more detail as follows.

图形用户界面, 图示

描述已自动生成

FIRST WEB-2.1 Metacritic

As mentioned in the introduction, Metacritic's official website will be the primary data source in our project as they can provide millions of data under different filters. Our goal is to build a Switch games database, we would like to catch the data about games that can play on Switch, and we will not set a time section. These selections will bring us to a web page that lists all Switch games under user scores from top to bottom. This web page will be the first web page we want to focus on under the first web crawler. Our primary goal in here is to collect the top 200 game names that stand as the primary key for the game table, which we will introduce in the Relational Database section. The web crawler workflow is shown in the following figure.

(Figure)

Before starting the work, we need to understand the URL of the web page. We found the difference in URLs from the first and second of the first web page. Since we want to collect the first 200 games in the first web scraping, this web page only contains 100 games on the first page. This will be a noticeable part when dealing with the programming code; we need a loop on it. As we have analyzed the web page, then we are ready to work. The first stage is to retrieve data. To archive this aim, we need to send a request to the target web page through an HTTP. We used a package called requests from the python library and included headers in the request. After completing this step, we will get a Response; here is the page's content to be obtained. The next stage aims to parse the content we just archived, done by using another package called BeautifulSoup. The content we get is in HTML format, and we parse with it and extract critical data such as game name, game ranking and the link to the second web through regular expressions. The final step is to save the data. We hold the data in CSV format locally.

So far to this step, we have done the data collection for the first web page. As we have the link to the second web page for each game, we are ready to work on the second web crawler, similar to working as the first web crawler. Firstly, we have to understand the structure of the HTML page. The data such as user score and the number of players are the main points we want to focus on. We used Xpath instead of regular expressions as we wanted to be more specific and accurate to the resulting data. The difference from the first web crawler is the way of saving data; this time, we would like to save it as xlsx form. Until this stage, we have collected data over the game name, game ranking, release date, scores, etc; 13 columns of relational data of the top 200 Switch games listed by the user score from Metacritic.

(Figure \_ column list)

* SECOND WEB
* YOUTUBE

1. **Data processing and Cleaning**

**-** FIRST WEB

* SECOND WEB
* YOUTUBE

1. **Data Storage**
2. **Relational Database**
3. **Exploratory Data Analysis**
4. **Database Value**
5. **Conclusion**
6. **Appendix**
7. **References**

<https://www.pocket-lint.com/games/reviews/nintendo/140007-nintendo-switch-review-console-specs-price-and-features>

<https://www.altexsoft.com/blog/engineering/what-is-api-definition-types-specifications-documentation/>