Under New Management

Managing and Manipulating Data

Charlie Chiccarine

PaTTAN Computer Science Praxis Prep

Databases

- Places to store data
- Many different flavors
  - Relational - SQL
  - Spreadsheet - Microsoft Excel
  - Personal - Microsoft Access

Uses for Databases

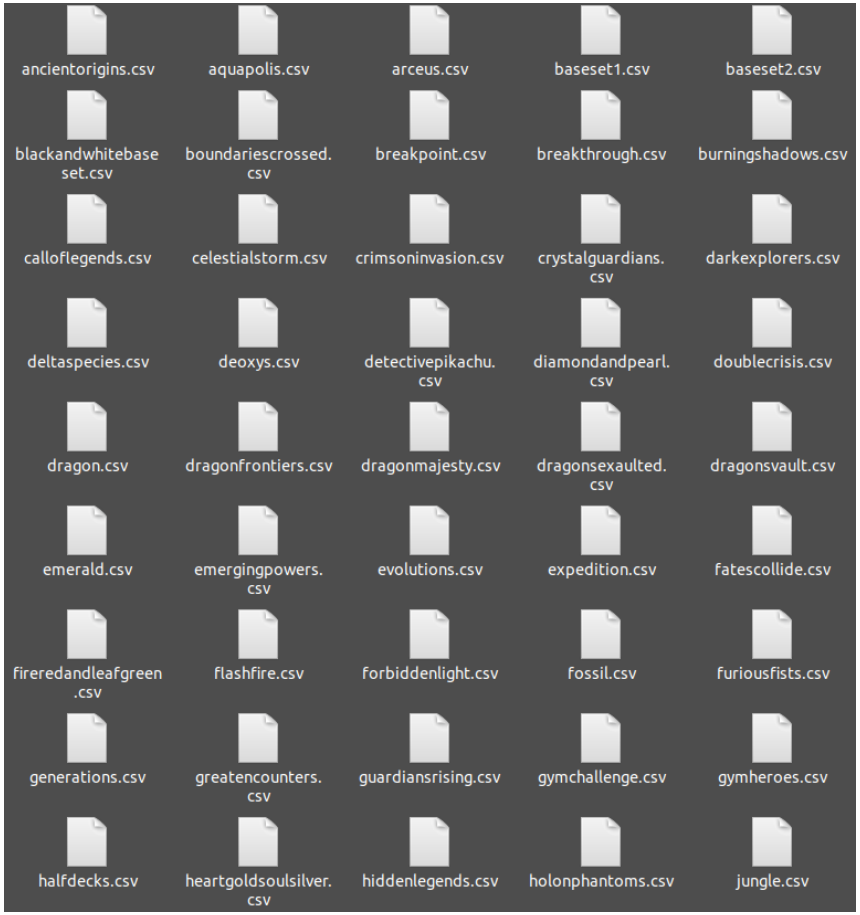
- Store large amounts of data
- In a comercial environment, use it to store large amounts of customers or transactions
- In a personal setting, use it to manage personal data like passwords

My Personal Database - Pokemon Cards

- I have 2670+ unique pokemon cards I collect
- Each card has lots of data I need to remember
- I made a database to store all the information on these files
- A collection of comma separated value (csv) files
- I have written a collection of Python scripts to manage the data

Pokemon Cards - Set Up

- I have a folder on my computer called "Pokemon_Files"
- In this folder I have 93 csv files
  - A csv file is a type of text file that organizes information in columns separated by commas
  - The first row is a header row of information about each column
- Each csv file has anywhere between 1 and 103 rows



Pokemon Cards - Set Up

- Let's take a closer look at these files
- As you can see the top row is the headers with each following row being a pokemon card
- A csv file has a similar set up as a spreadsheet

```
Pokedex,Pokemon,Type,Rarity,HoloFoil,Rarity Extra,Set No.,Set,Artist,Damaged,Price,Date,TCG Player Product ID
43,Oddish,G,1,,,1/98,,MAHOU,,,7/7/2019,XXXXXX
44,Gloom,G,2,,,2/98,,Masakazu Fukuda,,,7/30/2019,XXXXXX
182,Bellossom,G,2,Art,,4/98,,Mizue,,,,100616
167,Spinarak,G,1,,,5/98,,Naoyo Kimura,,,7/30/2019,XXXXXX
168,Ariados,G,2,,,6/98,,Hajime Kusajima,,,7/7/2019,XXXXXX
415,Combee,G,1,,,9/98,,Sumiyoshi Kizuki,,,7/30/2019,XXXXXX
416,Vespiquen,G,2,,,10/98,,kawayoo,,,6/15/2019,XXXXXX
136,Flareon,F,2,Reverse,,13/98,,sui,,,7/30/2019,XXXXXX
244,Entei,F,3,,,14/98,,Naoki Saito,,,8/6/2019,XXXXXX
637,Volcarona,F,3,,,18/98,,Kagemaru Himeno,,,7/30/2019,XXXXXX
129,Magikarp,W,1,,,19/98,,Akira Komayama,,,5/12/2019,101440
130,Gyarados,W,3,,,20/98,,Shin Nagasawa,,,5/12/2019,101441
134,Vaporeon,W,2,Reverse,,22/98,,kirisAki,,,7/7/2019,XXXXXX
181,M Ampharos EX,El,5,,,28/98,,5ban Graphics,,,7/7/2019,XXXXXX
201,Unown,P,1,Card,,30/98,,Akira Komayama,,,7/30/2019,XXXXXX
343,Baltoy,P,1,,Full Art,32/98,,Ayaka Yoshida,,,5/12/2019,101453
344,Claydol,P,3,Card,,33/98,,Satoshi Shirai,,,7/7/2019,XXXXXX
622,Golett,P,1,,,34/98,,Sanosuke Sakuma,,,5/12/2019,101457
194,Wooper,Fg,1,,,38/98,,Sumiyoshi Kizuki,,,5/12/2019,101460
195,Quagsire,Fg,1,,,39/98,,Naoyo Kimura,,,7/30/2019,XXXXXX
623,Golurk,Fg,1,,,41.98,,Suwama Chiaki,,,7/7/2019,XXXXXX
302,Sableye,D,2,,,44/98,,Hitoshi Ariga,,,5/12/2019,101464
686,Inkay,D,1,,,45/98,,Tomokazu Komiya,,,7/7/2019,XXXXXX
687,Malamar,D,1,,,46/98,,Masakazu Fukuda,,,6/15/2019,XXXXXX
374,Beldum,Sl,1,,,47/98,,Yuka Morii,,,7/7/2019,XXXXXX
280,Ralts,Fa,1,,,52/98,,Aya Kusube,,,5/12/2019,101474
281,Kirlia,Fa,2,,,53/98,,match,,,5/18/2019,101475
282,Gardevoir,Fa,3,Reverse,,54/98,,TOKIYA,,,5/18/2019,101476
546,Cottonee,Fa,1,,,55/98,,Kanako Eo,,,7/20/2019,XXXXXX
704,Goomy,Dr,1,,,58/98,,sui,,,7/7/2019,XXXXXX
705,Sliggoo,Dr,2,,,59/98,,Midori Harada,,,5/12/2019,101481
52,Meowth,N,1,,,61/98,,Kanako Eo,,,7/30/2019,XXXXXX
53,Persian,N,1,Full,,62/98,,Mitsuhiro Arita,,,6/15/2019,XXXXXX
T,Ace Trainer,Su,2,,,69/98,,Ken Sugimori,,,7/7/2019,XXXXXX
T,Ampharos Spirit Link,I,2,,,70/98,,5ban Graphics,,,7/30/2019,XXXXXX
T,Forest of Giant Plants,S,2,Card,,74/98,,Ryo Ueda,,,8/6/2019,XXXXXX
T,Hex Maniac,Su,2,,,75/98,,Yusuke Ohmura,,,7/30/2019,XXXXXX
T,Level Ball,I,2,Full,,76/98,,Toyste Beach,,,5/12/2019,101498
T,Lucky Helmet,I,2,,,77/98,,Toyste Beach,,,6/15/2019,XXXXXX
```

Pokemon Cards - Manipulating the Data

- How can we manipulate this plain text into data I can use?
- Since the data format is standard, I can manipulate it using scripts
  - There are always 13 columns that are always separated by commas
  - Column 2 is always the Pokemon name
- Python is a great language for scripting
  - A script is a smaller piece of code used for performing simple tasks
- Using Python, I can read a file line by line, and get information from it

Pokemon Cards - A Different Approach

- For me, this database set up is great and does exactly what I need it to
- However, it is very code heavy, so it's not the best option for most people

Pokemon Cards - Spreadsheet

- Let's load my csv file into Google Sheets
- To most people, this is easier to read
- Instead of running scripts to organize my data, I can just use the spreadsheet functions

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Pokedex	Pokemon	Type	Rarity	HoloFoil	Rarity Extra	Set No.	Set	Artist	Damaged	Price	Date	TCG Player Product ID
2	43	Oddish	G	1			1/98		MAHOU			7/7/2019	XXXXXX
3	44	Gloom	G	2			2/98		Masakazu Fukuda			7/30/2019	XXXXXX
4	182	Bellossom	G	2	Art		4/98		Mizue				100616
5	167	Spinarak	G	1			5/98		Naoyo Kimura			7/30/2019	XXXXXX
6	168	Ariados	G	2			6/98		Hajime Kusajima			7/7/2019	XXXXXX
7	415	Combee	G	1			9/98		Sumiyoshi Kizuki			7/30/2019	XXXXXX
8	416	Vespiquen	G	2			10/98		kawayoo			6/15/2019	XXXXXX
9	136	Flareon	F	2	Reverse		13/98		sui			7/30/2019	XXXXXX
10	244	Entei	F	3			14/98		Naoki Saito			8/6/2019	XXXXXX
11	637	Volcarona	F	3			18/98		Kagemaru Himeno			7/30/2019	XXXXXX
12	129	Magikarp	W	1			19/98		Akira Komayama			5/12/2019	101440
13	130	Gyarados	W	3			20/98		Shin Nagasawa			5/12/2019	101441
14	134	Vaporeon	W	2	Reverse		22/98		kirisAki			7/7/2019	XXXXXX
15	181	M Ampharos EX	El	5			28/98		5ban Graphics			7/7/2019	XXXXXX
16	201	Unown	P	1	Card		30/98		Akira Komayama			7/30/2019	XXXXXX
17	343	Baltoy	P	1		Full Art	32/98		Ayaka Yoshida			5/12/2019	101453
18	344	Claydol	P	3	Card		33/98		Satoshi Shirai			7/7/2019	XXXXXX
19	622	Golett	P	1			34/98		Sanosuke Sakuma			5/12/2019	101457
20	194	Wooper	Fg	1			38/98		Sumiyoshi Kizuki			5/12/2019	101460
21	195	Quagsire	Fg	1			39/98		Naoyo Kimura			7/30/2019	XXXXXX
22	623	Golurk	Fg	1			41/98		Suwama Chiaki			7/7/2019	XXXXXX
23	302	Sableye	D	2			44/98		Hitoshi Ariga			5/12/2019	101464
24	686	Inkay	D	1			45/98		Tomokazu Komiya			7/7/2019	XXXXXX
25	687	Malamar	D	1			46/98		Masakazu Fukuda			6/15/2019	XXXXXX
26	374	Beldum	Sl	1			47/98		Yuka Morii			7/7/2019	XXXXXX
27	280	Ralts	Fa	1			52/98		Aya Kusube			5/12/2019	101474
28	281	Kirlia	Fa	2			53/98		match			5/18/2019	101475
29	282	Gardevoir	Fa	3	Reverse		54/98		TOKIYA			5/18/2019	101476
30	546	Cottonee	Fa	1			55/98		Kanako Eo			7/20/2019	XXXXXX
31	704	Goomy	Dr	1			58/98		sui			7/7/2019	XXXXXX
32	705	Sliggoo	Dr	2			59/98		Midori Harada			5/12/2019	101481
33	52	Meowth	N	1			61/98		Kanako Eo			7/30/2019	XXXXXX
34	53	Persian	N	1	Full		62/98		Mitsuhiro Arita			6/15/2019	XXXXXX

Scale and Purpose

- Why did I chose to use csv files and Python scripts instead of a spreadsheet?
- Scope, scale, purpose, and applicability
- Scope and scale deal with how large the project is, I don't want my database to be super big
- Purpose and applicability deal with how applicable the design is, I am using my data in a very code heavy way. It would not be applicable to use a spreadsheet for my purposes
- I chose to use csv files and Python scripts because for my purposes, it is faster

Backup Data

- Since my database is fairly small, I can create multiple back ups
- A back up is a copy of the database in case I accidentally delete or alter the data in a wrong way
- I also use git, which is a type of version control, to make sure I don't make any changes I don't want to undo
- I use the GitHub servers to store all my csvFiles, since all my data is on a server, I can go back to it as long as I can connect

Recap

- Databases store large amounts of data
- There are many different types of databases
- Different types of databases are better for different purposes
- Back up your files to keep them safe