Data Skills Bootcamp Portfolio

Matthew Logue

Contents

My Background

I am a recent physics graduate with a 2:1 degree from the University of Birmingham and experience working as a software developer. I have strong data analysis, communication, and collaboration skills gained from my degree, in addition to database management skills from my role as a developer. I am looking to secure a position which will allow me to utilise these skills to solve real-world problems in a data-driven business.

In order to develop my data visualisation skills I have taken part in a level 3 skills bootcamp run by Cambridge Spark, during which I have created this portfolio to demonstrate the skills I have learned.

Project Aims

- I have analysed data relating to characters that players create for the role-playing game Dungeons and Dragons.
- The data cover characters created between April 2018 and November 2022 and are collected from two different character creation apps:
  - D&DBeyond (official app)
  - Fifth Edition Character Sheet (unofficial app)
- The goal of the project is to determine which character options are most popular and which factors determine that popularity.

Data Overview

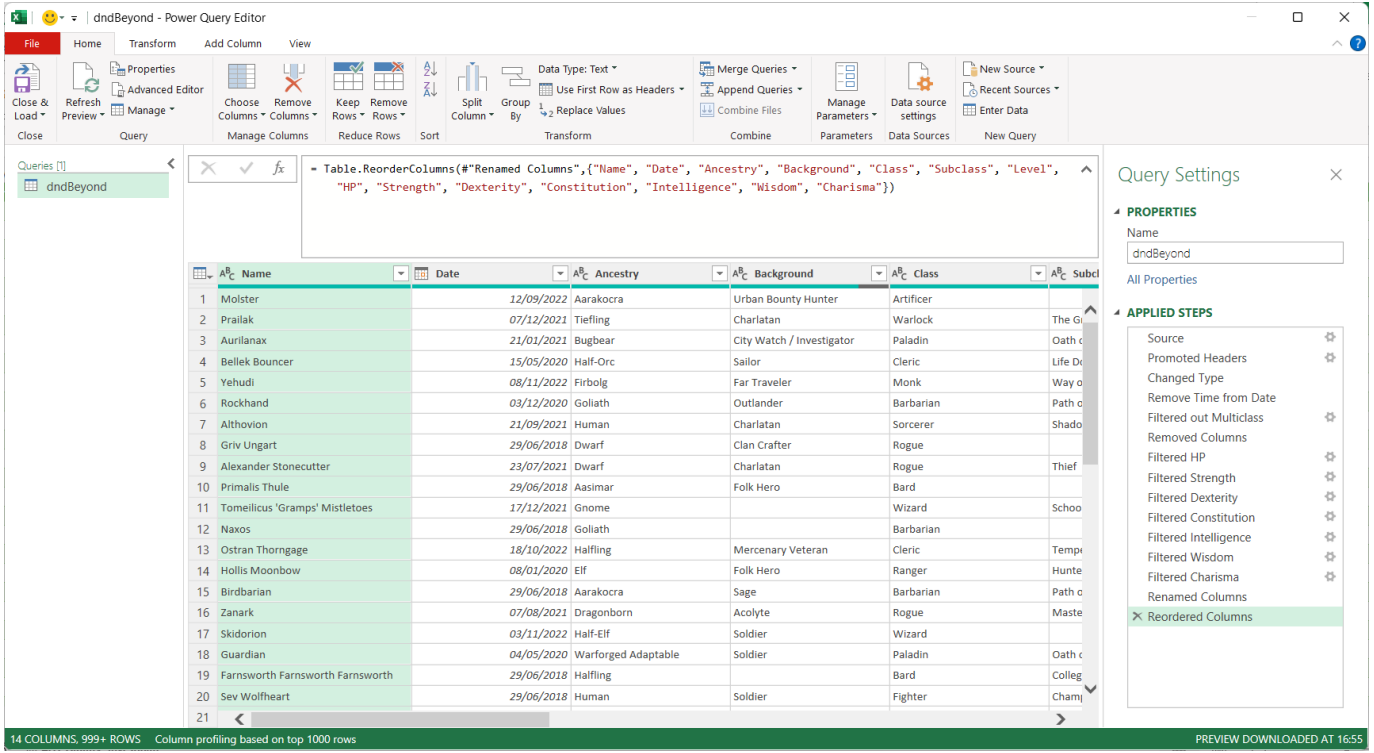
- Once cleaned, the data set comprises around 400,000 records obtained from the official character builder, with an additional 7,500 records from the other app.
- The date is the date that the character was last modified; and the country is the location of the user.
- The ancestry, sub-ancestry, background, class, subclass, and alignment are all chosen by a player from a set list of options.
- The rest of the fields are various measures of the capability of a character.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	CharacterID	Date	Ancestry	Sub-ancestry	Background	Class	Subclass	Level	HP	AC	Strength	Dexterity	Constitution	Intelligence	Wisdom	Charisma	Alignment	Country
51	49	17/04/2018	Elf	Wood	Urban Bounty Hunter	Rogue	Scout	4	23	14	10	17	10	10	15	14	Chaotic Neutral	United States
52	50	17/04/2018	Tiefling		Haunted One	Warlock	Great Old One	1	10	12	8	13	14	13	10	17		United States
53	51	17/04/2018	Elf	Wood	Criminal	Rogue		1	10	15	12	20	15	15	12	13		United States
54	52	18/04/2018	Human		Soldier	Rogue	Assassin	3	24	15	9	16	14	14	14	11	Neutral	United States
55	53	18/04/2018	Elf	High	Haunted One	Wizard		1	8	13	9	16	14	16	10	9		United States
56	54	19/04/2018	Aasimar	Fallen	Homebrew	Druid	Dreams	6	45	16	10	14	14	10	16	12	Neutral Good	United States
57	55	19/04/2018	Half-Elf		Acolyte	Paladin		2	24	18	10	14	14	10	14	16		Brazil
58	56	19/04/2018	Elf	Wood	Hermit	Rogue		2	18	14	10	16	13	12	14	12		Brazil
59	57	19/04/2018	Tabaxi		Sailor	Bard	Swords	3	27	15	12	18	17	13	14	18		Brazil
61	59	19/04/2018	Gnome	Rock	Faction Agent	Blood Hunter	Mutant	10	84	15	13	16	14	14	12	12		United States
62	60	20/04/2018	Elf	Drow	Urchin	Rogue		1	8	14	13	17	12	14	10	16		Peru
63	61	21/04/2018	Human		Hermit	Monk	Open Hand	5	45	17	9	18	14	11	16	13	Lawful Good	United States
64	62	21/04/2018	Elf	High	Noble	Rogue		1	8	13	8	17	10	15	12	13	Lawful Good	United States
65	63	21/04/2018	Half-Elf		Faction Agent	Sorcerer	Draconic Bloodline	20	182	17	14	18	18	15	18	24		Brazil

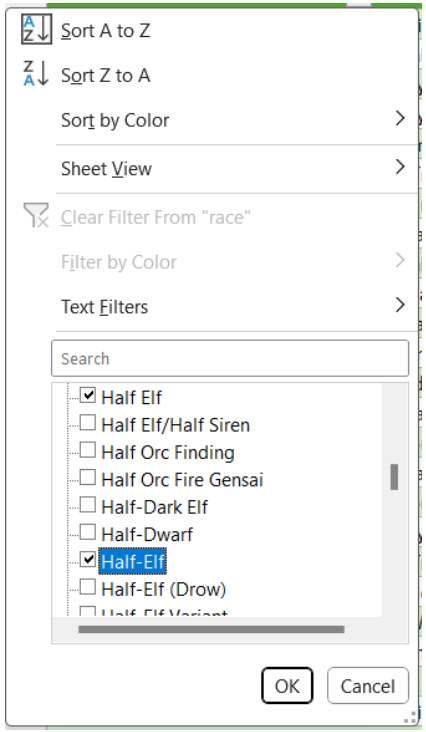
Methodology - Cleaning

I initially used Power Query to select only the relevant columns from each data set, and to filter out invalid values.

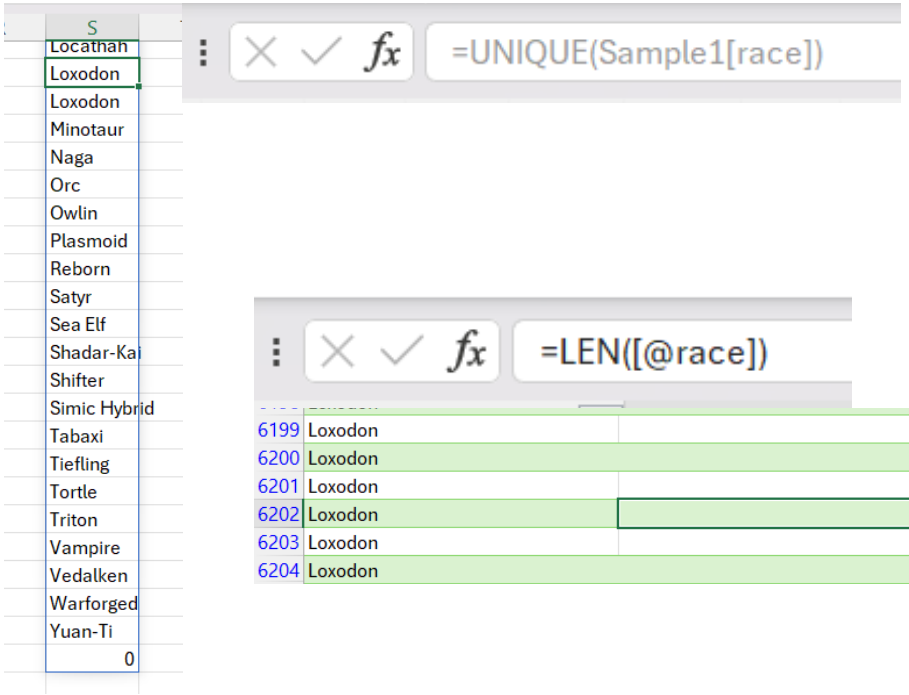
I also renamed and reordered the columns to be consistent between the two data sets.



Methodology - Cleaning



After loading the data from power query into an excel worksheet, I had to inspect the contents of each column to check for inconsistencies e.g. “Half Elf” and “Half-Elf” or “Loxodon” and “ Loxodon”.



I used the UNIQUE and LEN functions to confirm that I had found each variation of the valid values.

Methodology - Cleaning

I also split the data in the “Ancestry” and “Class and Level” fields for easier analysis later.

102	Air Genasi
103	Air Genasi
104	Air Genasi
105	Air Genasi
106	Air Genasi
107	Air Genasi
108	Air Genasi
678	Earth Genasi
679	Earth Genasi
680	Earth Genasi

102	Genasi	Air
103	Genasi	Air
104	Genasi	Air
105	Genasi	Air
106	Genasi	Air
107	Genasi	Air
108	Genasi	Air
678	Genasi	Earth
679	Genasi	Earth
680	Genasi	Earth

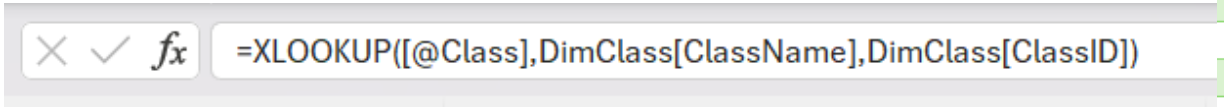
Font Alignment Number Styles

=LEFT([@[Class Level]],LEN([@[Class Level]])-LEN([@Level])-1)

B	C	D	E	F	G
estry	Sub-ancestry	Background	Class Level	Level	Class
ıasi	Air	Acolyte	Warlock 4	4	=LEFT([@[Class Level]],L
ter	Beasthide	Acolyte	Rogue 14	14	R LEFT(**text**, [num_chars])
ling	Bloodline of Dispater	Acolyte	Warlock 2	2	Warlock
ling	Bloodline of Fierna	Acolyte	Paladin 1	1	Paladin
ling	Bloodline of Levistus	Acolyte	Bard 5	5	Bard
ling	Bloodline of Levistus	Acolyte	Paladin 2	2	Paladin
ling	Bloodline of Mephistopheles	Acolyte	Paladin 2	2	Paladin
ling	Bloodline of Mephistopheles	Acolyte	Warlock 8	8	Warlock
ling	Bloodline of Zariel	Acolyte	Barbarian 2	2	Barbarian
ling	Bloodline of Zariel	Acolyte	Paladin 5	5	Paladin
ling	Bloodline of Zariel	Acolyte	Paladin 5	5	Paladin
ling	Bloodline of Zariel	Acolyte	Paladin 1	1	Paladin

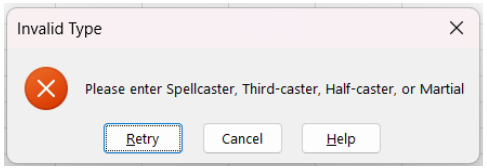
Methodology - Normalisation

Once I had cleaned the data, I normalised it by moving the values for categorical fields into their own tables and referencing them by ID number. This allowed additional information about each value to be stored more efficiently.



Class	ClassID
Warlock	12
Paladin	8
Cleric	4
Monk	7
Barbarian	1
Wizard	13
Sorcerer	11
Fighter	6
Rogue	10
Warlock	12
Cleric	4
Ranger	9
Barbarian	1
Rogue	10
Rogue	10
Paladin	8
Fighter	6

	ClassID	ClassName	HitDice	NumberSkills	Type
2	0	Artificer	8	2	Spellcaster
3	1	Barbarian	12	2	Martial
4	2	Bard	8	3	Spellcaster
5	3	Blood Hunter	10	3	Third-caster
6	4	Cleric	8	2	Spellcaster
7	5	Druid	8	2	Spellcaster
8	6	Fighter	10	2	Third-caster
9	7	Homebrew			Other
10	8	Monk	8	2	Martial
11	9	Mystic	8	2	Spellcaster
12	10	Paladin	10	2	Half-caster
13	11	Ranger	10	3	Half-caster
14	12	Rogue	8	4	Third-caster
15	13	Sorcerer	6	2	Spellcaster
16	14	Warlock	8	2	Spellcaster
17	15	Wizard	6	2	Spellcaster

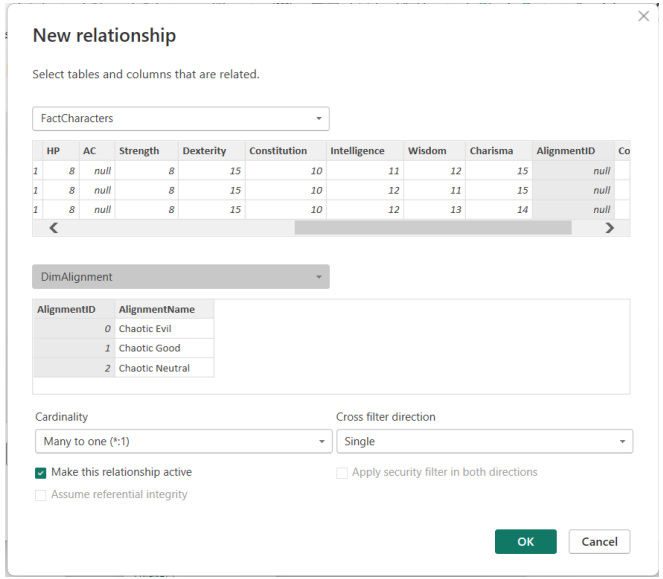


I also added data validation to make it easier to extend the table if new classes are created.

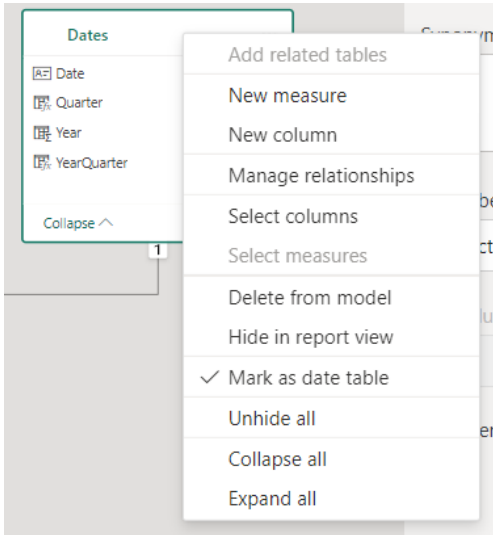
Methodology – Power BI

After cleaning the data in Excel, I used Power Query to append the two data sets and import them into Power BI for visualisation.

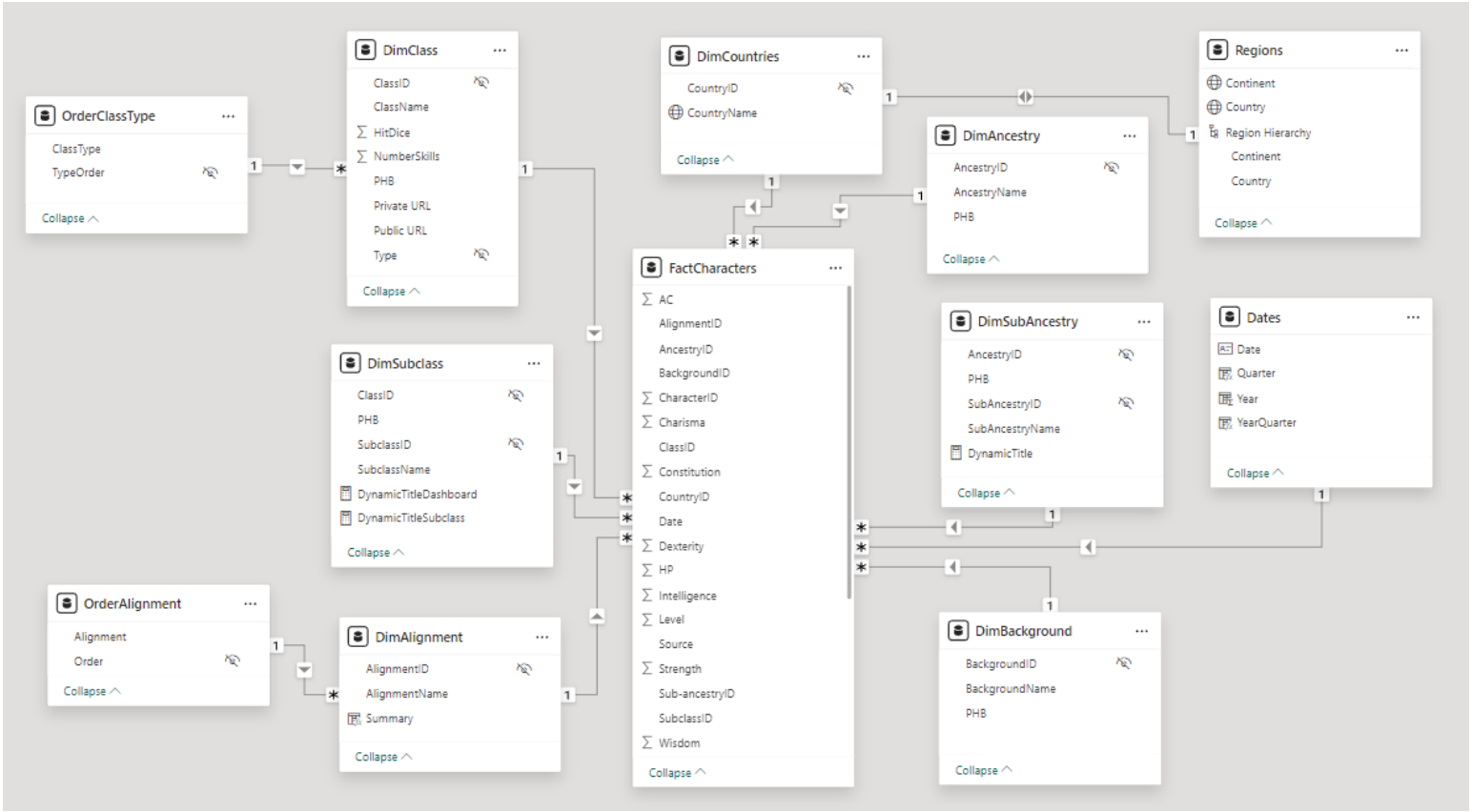
To ensure that the visuals would function correctly, I had to create relationships and DAX measures, and add a date table to the data model.




```
% of Total = DIVIDE(
    COUNT('FactCharacters'[CharacterID]),
    CALCULATE(
        COUNT('FactCharacters'[CharacterID]),
        ALL()
    )
)
```

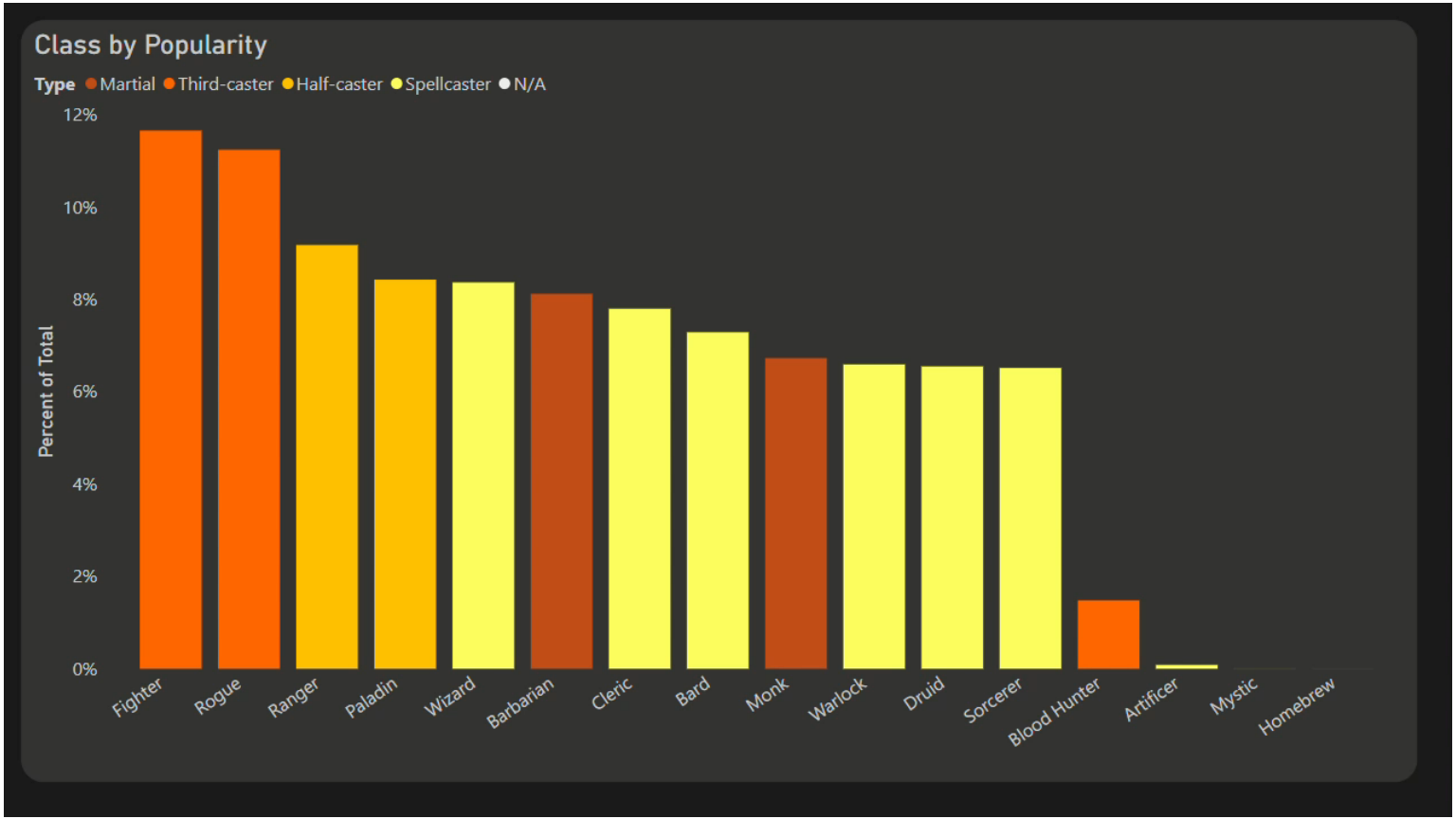


Medodology – Final Data Model

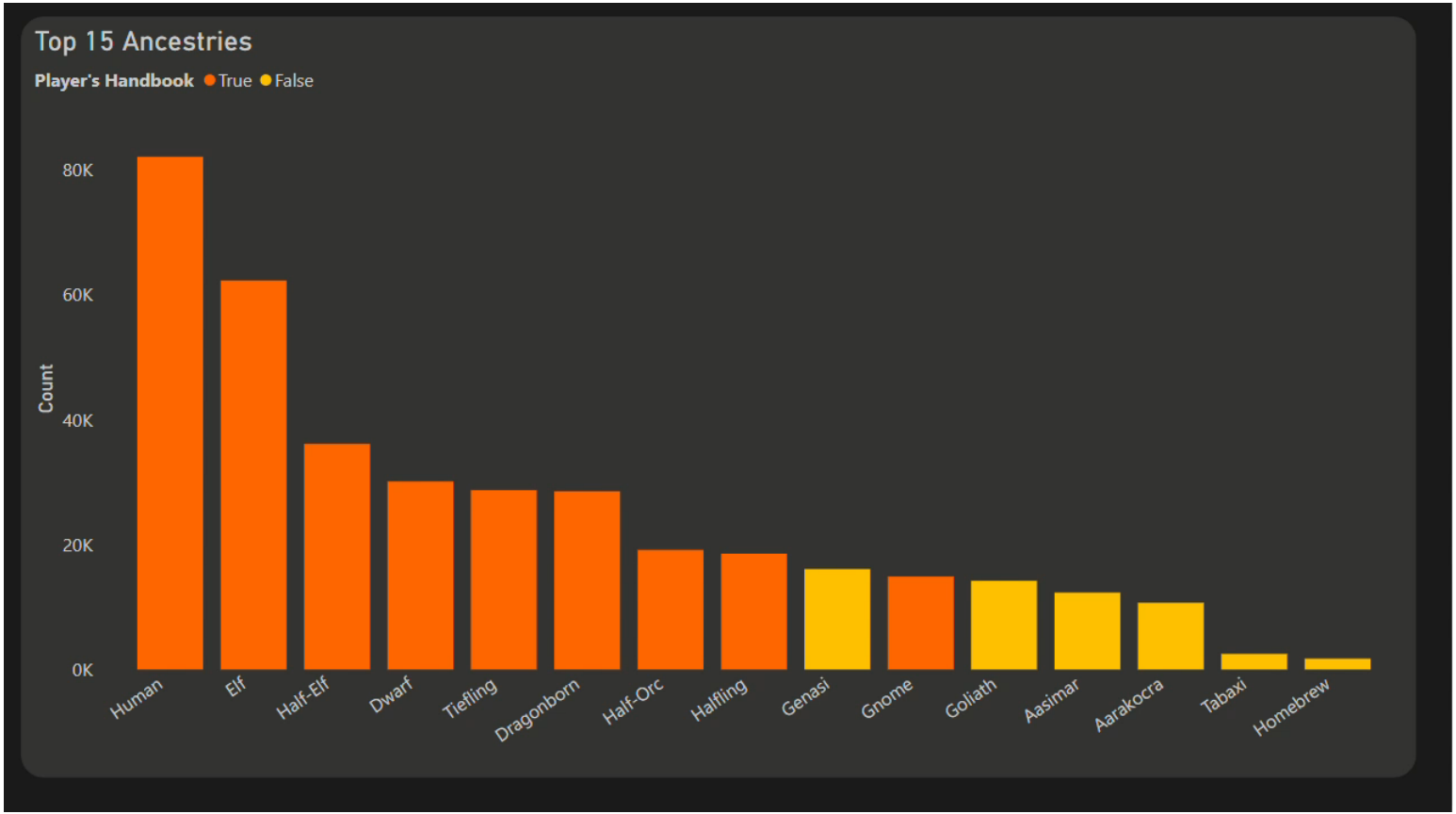


Results - Dashboard

Results – Classes



Results – Ancestries



Analysis

- Most characters are created using a small number of options.
  - 60% of characters only use options that were released in the first source book published in 2014.
  - Out of 98 possible backgrounds, 84% of characters use the top 7, of which 6 are from the first source book.
  - With 167 subclasses available across 15 classes, 70% of characters use the most popular subclass for each class.
- Most characters have a chaotic or good alignment, with only 6% of characters being evil, as expected for a heroic fantasy game.
- Most users are in the United States or United Kingdom reflecting the lack of availability of translations of source books.

Further Research and Links

A GitHub repository for the project can be found here. It contains the cleaned data set, along with the Power BI file I have produced.

Further avenues for exploration of the data could include:

- Analysis of more granular choices such as individual skills and spells learned by a character.
- Since the data from the unofficial app is such a small part of the data set, it hasn't been possible to reliably determine whether there is a difference between the two apps.
- It would also be interesting to collect data from people playing the game offline to determine whether the same trends can be seen.

Thank you!