PROJECT PROPOSAL OF MAGIX CONSULTANTS [TEAM 22]

Project Overview

This is a school management database developed for the Hogwarts School of Magic by our team of analysts from Magix Consultants. We are situated in Diagon Alley and have been serving the wizarding community since the inception of the Gringotts bank. We appreciate the trust that wizarding schools throughout the world have placed upon us all these years.

We have taken up a data driven approach to find the bottlenecks in the school functionality, areas of growth and challenges faced by staff and students alike. We feel that SQL databasing can help Hogwarts' position in the future and will be able to get a big picture view of the school without being bogged down into the tactical day-to-day operations.

Business Problem

Until now, the school has been maintaining most of its records on paper and the management has approached Magix Consultants for a solution to switch to cloud solutions so that they don't get destroyed during Voldemort's continuous attacks on Hogwarts.

Also another main purpose of creating this school management database is to make the activities of the whole school more streamlined and to optimize the record-keeping capabilities. Our system will emulate the whole processes that make up the day-to-day tasks and provide essential data support to its stakeholders so that they can be more efficient in their respective roles. The implementation of the system was done using MySQL workbench, with visualizations from Tableau, Python and Rstudio, to help the management at school understand how the database was implemented and the business concerns were addressed.

Domain Logic/Database Architecture/Business Logic

We have built this database around three core services, namely classes, students and employees. In order to maintain precision, spells and potions, which are the most valued skills for seeking internships/jobs, were defined in a separate table. Similarly, for referencing purposes, we decided to maintain a different table for houses and grades, instead of adding them to students' table.

In order to maintain ease-of-use, organizational policies and to eliminate the necessity of a Business Logic Layer, we have built or provided avenues to contain all the business logic within the database itself, with the help of stored functions and procedures. Therefore, we aim to eliminate any dependence on external user interfaces for the purposes of database management.

Data Cleaning/Data

Our analysts faced some challenges while putting together the database that demanded data transformations. The key data transformation was imputing the missing values, since the data was stored in flying books where the ink has faded with time. The consulting team first tried to calculate

the average of the missing values, replaced them with zero or completely removed the missing entries.. Next we encountered the issue of inconsistent recorded values, some of the values were more than the 3 sigma from mean. Such records were removed or suitably modified. Lastly, we noticed that some of the values were too long to be captured in database column type, so we trimmed the values to fit into the preferred data length, to save on the storage costs of the database.

Table Creation

The table creation process of relatively simple. We identified all the essential attributes within each data structure, along with an unique identifier to be the primary key.

'Class' table contains the class name, subject ID, professor, credit, location, day and time the class is offered.

'Class level' table has just about enough information about the different class levels/grade levels that students change every year.

'Student' table has all the important information about each and every student at Hogwarts. This is one of the most important tables in the database because it is linked to most of the tables.

The `house` table describes the house name, element, house colors, patron ghost, head of the house and founder of the house.

'Employee' table has brief details like their name, gender and type related to the professors and staff working at Hogwarts.

'Potion' table has all the potions used in all of Potions classes currently taught by Professor Snape.

'Spell' table has all the spells used by all the professors taught and learned by students at Hogwarts. It is linked to classes to understand where they all belong.

'House' table has all the information of the four magnificent houses created at Hogwarts and linked to students.

'Grade' table describes the level of grades awarded to students.

'Pet' table has information on all the different types of animals that Hogwarts student own but however are not linked to students or any other owners.

'Club' table includes all the clubs running in Hogwarts currently.

Relevant Attributes/Instances

We have outlined a few of the important attributes/instances found in our database below:

- Students can always look up the relevant spells they need to learn in advance from the classes they would like to enroll in.
- `ClassId` column was modeled from the `subjectID` column, so as to maintain the ease of use for students wishing to sign up for their classes.

- To facilitate better connectivity, owner names in the `Pet` table have been split into 3 attributes, namely `ownerFirstName`, `ownerMiddleName` and `ownerLastName`

Relationships (between tables)

As our database is modelled based on the three core services of classes, students and employees, our database relationships are mostly grouped on the same grounds. The `Class` table is at the core of class part of the database and shares many-to-one relationships with the `Employee` and `Subjects` tables, which allows the user to reference the professor and subject of each class. The `Potion` and `Spell` tables relate to `class` table in a one-to-many relationship, as there can be multiple spells/potions taught in one class. The other sub-group was allocated to the services for students. The `House`, `Club` and `Grade` tables are all joined in a one-to-many relationships with the `Student` tables, as they concern features used only by students. In order to link both sub-groups, we have used the `ClassLevel` table, which has a many-to-one relations to both `Student` and `Class` tables, as it is the only service which decides which classes can be taken by which student. Finally, the pet table is the only table that is not connected to any of the tables in the database because it was hard to link up the owners from different tables.

Normalization

Some of the old data that was missing was found from the old library archives at Hogwarts. Librarian Pince has been extremely helpful in this regard. Some basic assumptions were made regarding the structure of certain tables.

In order to satisfy 1NF, all instances with a secondary result were divided into two different rows. This can be clearly seen in the ‘employees’ table, where an employee with two distinct roles was mentioned twice, with a unique employee number assigned to each position. We also extracted the ‘subjects’ from the ‘classes’ table in order to maintain its exclusivity and a foreign key column was added to the ‘Subjects’ table, which refers to the primary key of the ‘Classes’ Table. Therefore, the use of non-simple domains was reduced, and the table was changed from its un-normalized form to 1NF.

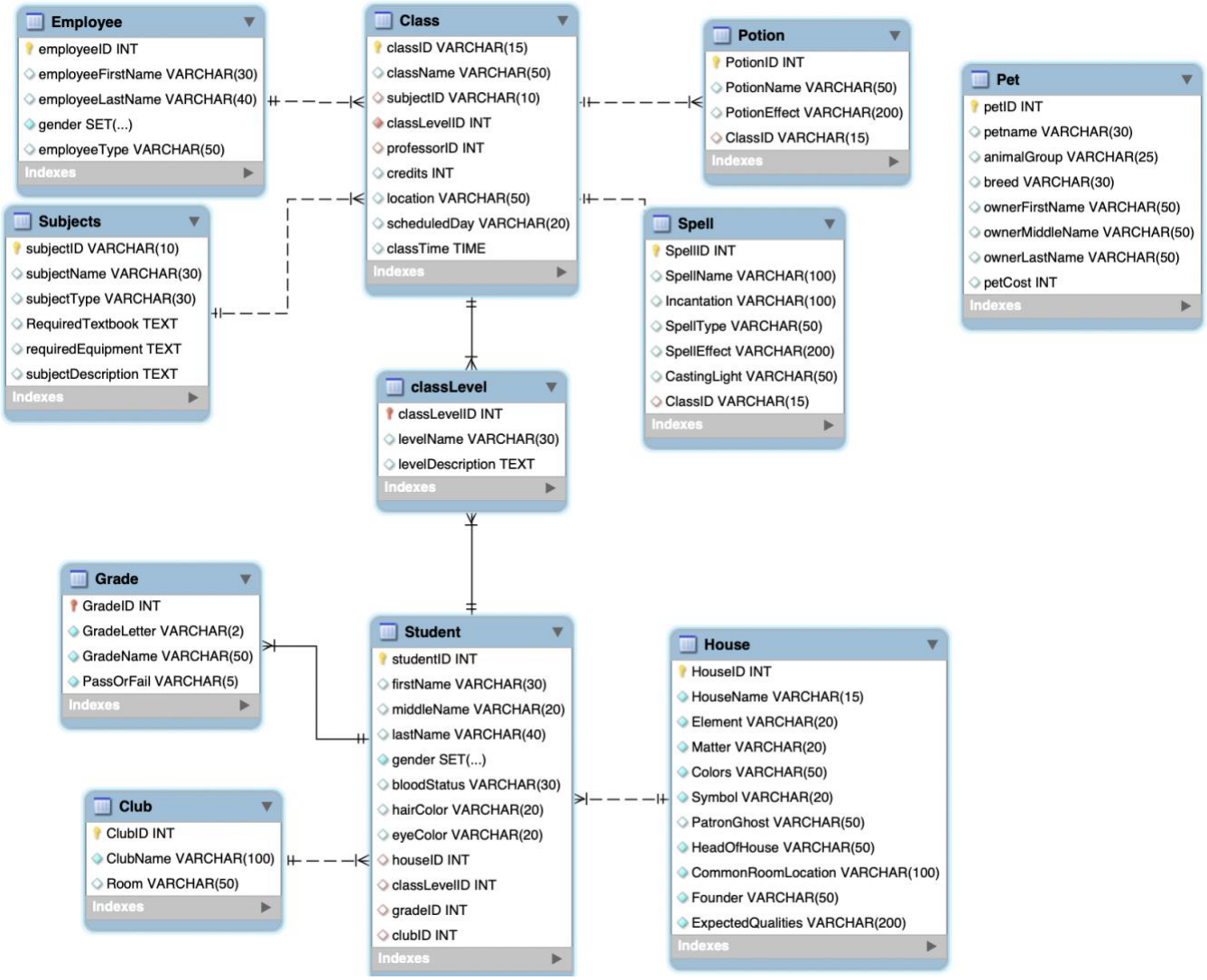
Due to the consistent use of unique identifiers for each instance, all tables in our database system conform to 2NF. Furthermore, all tables also conform to 3NF, as there is no transitive functional dependency between any of them.

Conclusion

Magix Consultants hopes to receive good feedback from the school about the database we have built for their effective management. Furthermore, the headmaster is so impressed with the scalability of SQL, that we hope to eventually help expand the database management and run it on an off-site server with multiple redundancies, so that dark magic will not be able to destroy the Hogwarts.

ER Model screenshot

Our ER Model was reverse engineered from our database in the mySQLWorkbench application.



Business Questions, with corresponding result screenshots

Business Question 1: The headmaster thinks that they need to revise the school syllabus and wants to check how many subject types are currently available at Hogwarts. After finding the current buckets, he will use this information to decide whether to add '*Muggle Language*' as an Elective or Extra-Curricular.

```
SELECT SubjectType,COUNT(SubjectType) AS `Number of Subjects` FROM Subjects
    GROUP BY SubjectType
    ORDER BY `Number of Subjects`;
```

Result Grid

SubjectType	Number of Subjec...
Extra-curricular	8
Core	8
Elective	9

Result 1

Business Question 2: The head of Slytherin wants to find all its member students who have the blood status of 'Muggle-born', to put in his final house diversity report.

```
SELECT studentID, concat(firstName," ",middleName," ",lastName) AS `Name of Student` FROM Student
    WHERE bloodStatus = 'Muggle-born' and houseID = 4;
```

Result Grid

studentID	Name of Student

Result 2

Business Question 3: The school will be voting to allocate a separate building for housing student pets and wants to know how many pets are owned by students and professors respectively. Knowing the number of such pets will help determine the size of the building.

```
SELECT IF(S.StudentID IS NULL , 'Student', 'Professor') AS `Owner Type`, count(petID) AS `Number of Pets`
FROM Pet as P
LEFT JOIN Student AS S ON P.OwnerFirstName = S.FirstName AND P.OwnerLastName = S.LastName
GROUP BY `Owner Type`;
```

Result Grid

Owner Type	Number of Pets
Student	11
Professor	7

Business Question 4: Some cauldrons have exploded, and the school wants to know which classes use a cauldron in order to restock them. This helps to keep track of the demand for equipment.

```
36 •  SELECT C.ClassName, S.SubjectName
37    FROM Class AS C
38    JOIN Subjects AS S ON C.SubjectID = S.SubjectID
39    WHERE S.requiredEquipment LIKE '%cauldron%';
40
```

100% 55:32

Result Grid Filter Rows: Search Export:

ClassName	SubjectName
Alchemy I	Alchemy
Alchemy III	Alchemy
Alchemy III	Alchemy
Alchemy IV	Alchemy
Alchemy V	Alchemy
Potions I	Potions
Potions II	Potions
Potions III	Potions
Potions IV	Potions
Potions V	Potions

Business Question 5: The School wants to report the level ranking within each level to its students, so that the students can see how they compare to their peers.

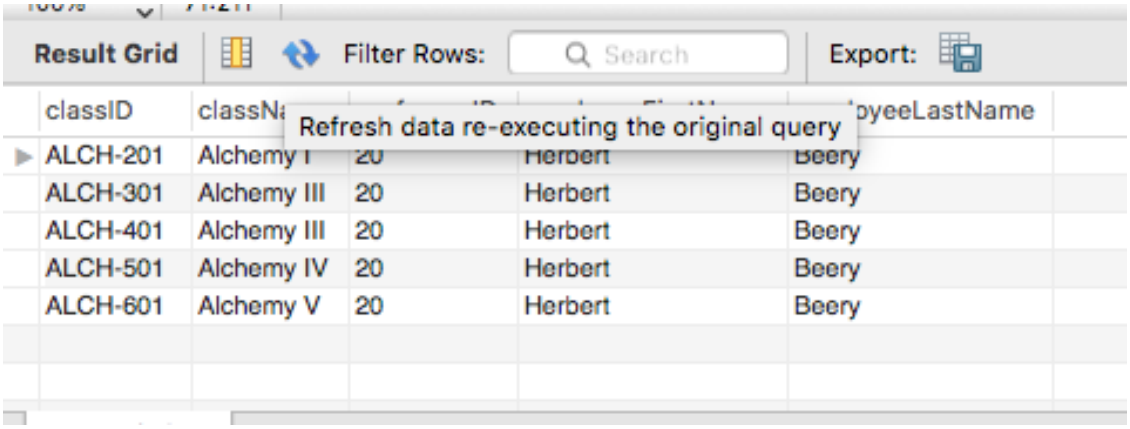
```
45 •  SELECT concat(S.firstName," ",S.middleName," ",S.lastName) AS `Name of Student` , CL.levelName, G.GradeName,
46              RANK() OVER (PARTITION BY S.ClassLevelID ORDER BY S.GradeID,S.HouseID) AS `Level Rank`
47    FROM Student AS S
48    JOIN ClassLevel AS CL ON S.ClassLevelID=CL.ClassLevelID
49    JOIN Grade AS G ON S.GradeID= G.GradeID;
```

100% 38:46

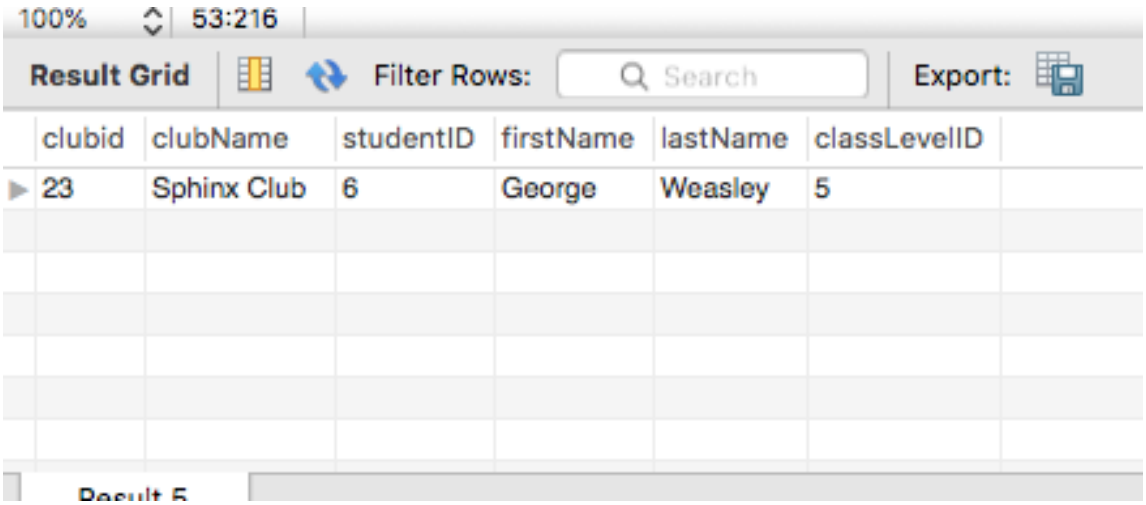
Result Grid Filter Rows: Search Export:

Name of Student	levelName	GradeName	Level Rank
Ronald Bilius Weasley	First Year	Outstanding	1
Marietta Edgecombe	First Year	Outstanding	2
Rose Granger-Weasley	First Year	Exceeds Expectations	3
Cho Chang	First Year	Exceeds Expectations	4
Graham Montague	First Year	Exceeds Expectations	5
Lee Jordan	First Year	Acceptable	6
Angelina Johnson	First Year	Acceptable	6
Tom Marvolo Riddle	First Year	Acceptable	8
Harry James Potter	First Year	Poor	9
Colin Creevey	First Year	Poor	9
Zacharias Smith	First Year	Dreadful	11
Hermione Jean Granger	First Year	Troll	12
Oliver Wood	Second…	Outstanding	1
Michael Corner	Second…	Outstanding	2
Vincent Crabbe	Second…	Outstanding	3
Scorpius Hyperion Mal…	Second…	Outstanding	3
Dean Thomas	Second…	Poor	5
Blaise Zabini	Second…	Poor	6
Penelope Clearwater	Second…	Dreadful	7
Cormac McLaggen	Second…	Troll	8
Dennis Creevey	Second…	Troll	8

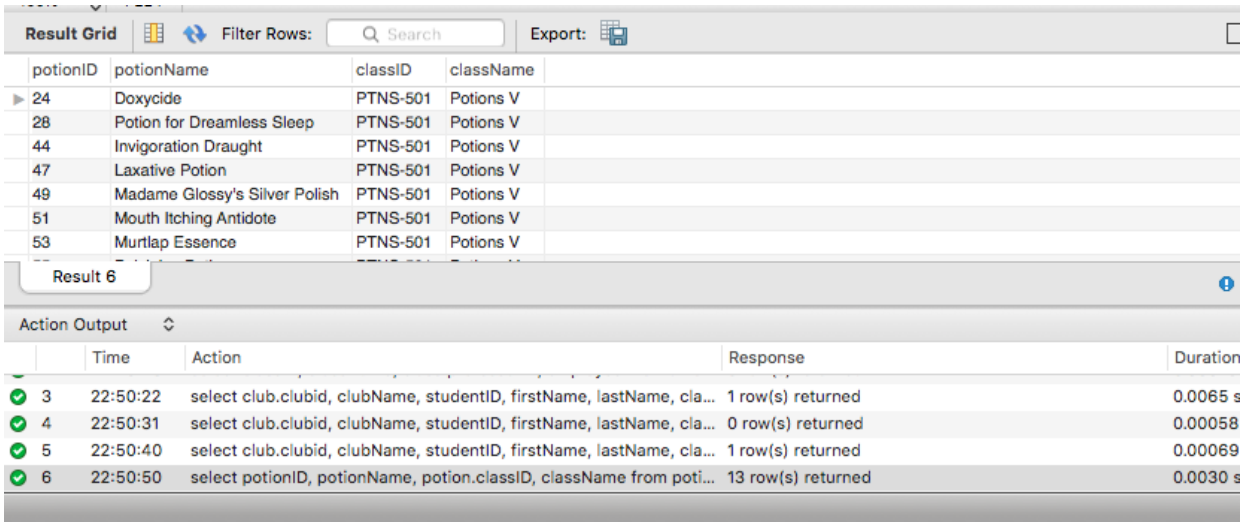
Business Question 6: Dumbledore wants to figure out the salary to give to the new professor of Alchemy as Professor Beery is transferring to another wizarding school in Greece. For this, we would like to find out the number of classes that Professor Beery currently teaches.



Business Question 7: In order to maintain funding throughout the academic year, SpinxClub needs students from all class levels. Find out which class levels are missing from the club.



Business Question 8: To teach the fifth year, Professor Snape needs to stock up on some potions that have been secretly missing from his storeroom. Find out how many potions are required for this.



Result Grid | Filter Rows: Search | Export:

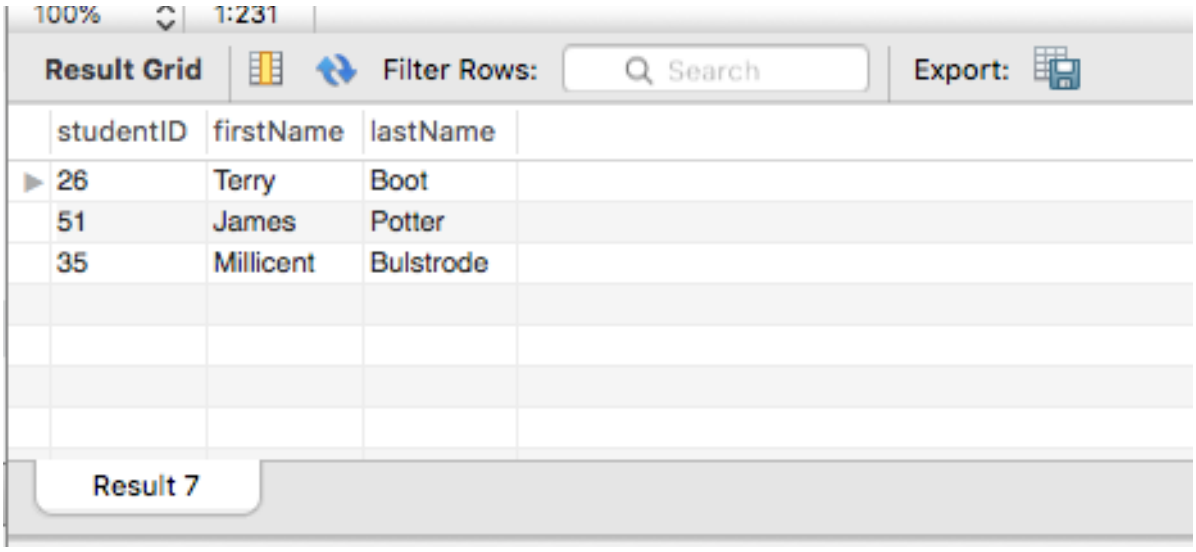
potionID	potionName	classID	className
24	Doxycide	PTNS-501	Potions V
28	Potion for Dreamless Sleep	PTNS-501	Potions V
44	Invigoration Draught	PTNS-501	Potions V
47	Laxative Potion	PTNS-501	Potions V
49	Madame Glossy's Silver Polish	PTNS-501	Potions V
51	Mouth Itching Antidote	PTNS-501	Potions V
53	Murtlap Essence	PTNS-501	Potions V

Result 6

Action Output

	Time	Action	Response	Duration
3	22:50:22	select club.clubid, clubName, studentID, firstName, lastName, cla...	1 row(s) returned	0.0065 s
4	22:50:31	select club.clubid, clubName, studentID, firstName, lastName, cla...	0 row(s) returned	0.00058
5	22:50:40	select club.clubid, clubName, studentID, firstName, lastName, cla...	1 row(s) returned	0.00069
6	22:50:50	select potionID, potionName, potion.classID, className from poti...	13 row(s) returned	0.0030 s

Business Question 9: The Ministry of Magic would like to offer internships to interested students from sixth year who have performed outstandingly and exceeded expectations in their O.W.L.S.



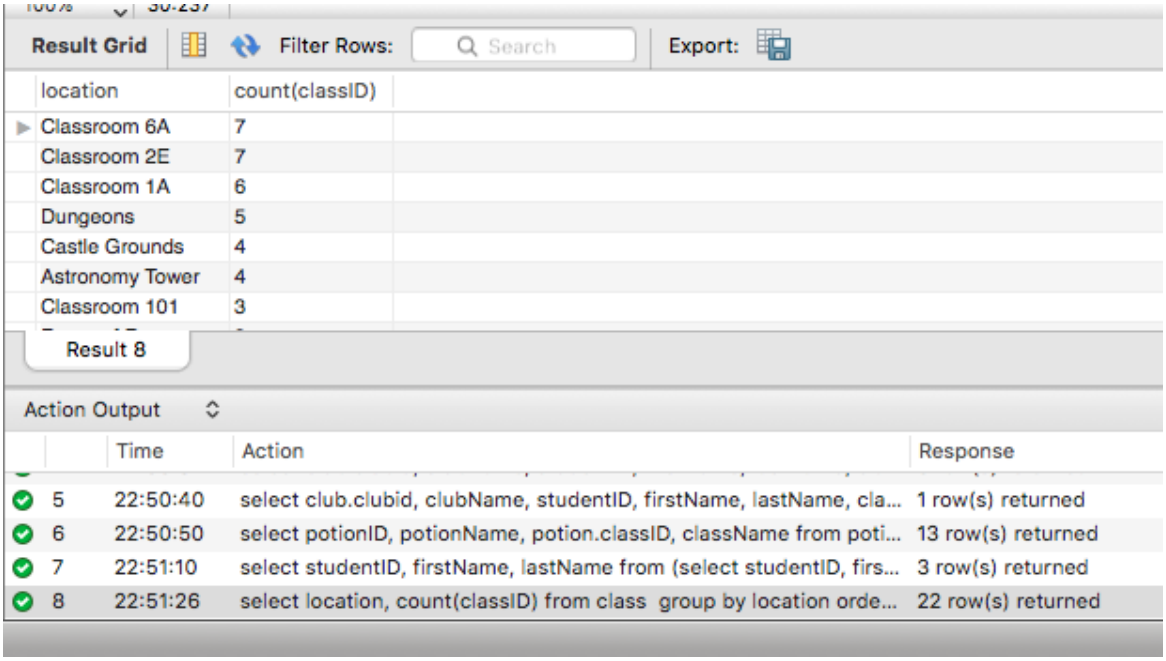
100% 1:231

Result Grid | Filter Rows: Search | Export:

studentID	firstName	lastName
26	Terry	Boot
51	James	Potter
35	Millicent	Bulstrode

Result 7

Business Question 10: The school management would like to renovate the classroom locations that hold most classes. Better interior solutions will help students focus better on learning.



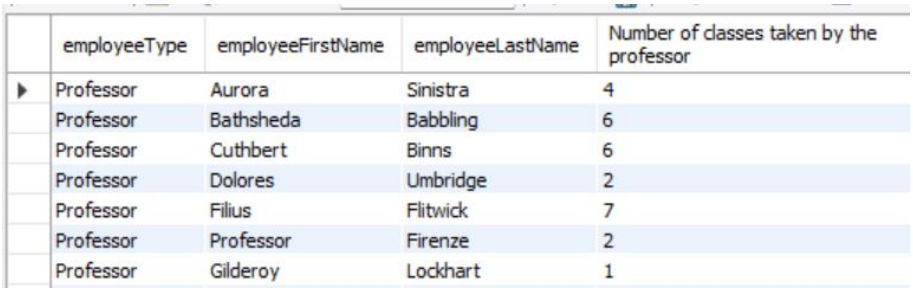
Business Question 11: Professor Minerva McGonagall with id 22 is interested in the number of credits she is doing over a particular week.

professorID	scheduledDay	sum(credits)
22	Tuesday	6
22	Friday	3
22	Monday	6
22	Saturday	3
22	Thursday	3

Business Question 12: 3 friends who got their grade letters in the result post wanted to know whether they passed or failed.

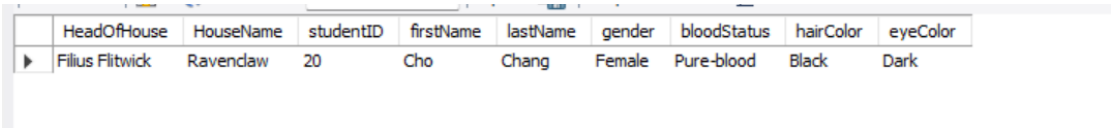
GradeLetter	PassOrFail
E	Pass
P	Fail
T	Fail

Business Question 13: The administration is interested in knowing the number of classes taken by each professor this semester.



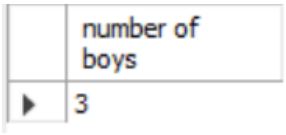
employeeType	employeeFirstName	employeeLastName	Number of classes taken by the professor
Professor	Aurora	Sinistra	4
Professor	Bathsheda	Babbling	6
Professor	Cuthbert	Binns	6
Professor	Dolores	Umbridge	2
Professor	Filius	Flitwick	7
Professor	Professor	Firenze	2
Professor	Gilderoy	Lockhart	1

Business Question 14: Head of the Ravenclaw house, Professor Filius Flitwick is interested in finding a girl, who doesn't have a middle name and is pure blooded in their house for the quidditch tournament.



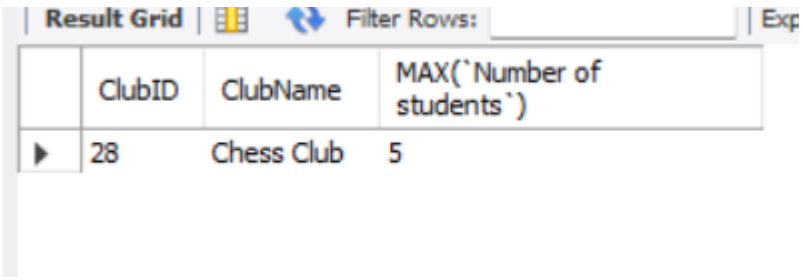
HeadOfHouse	HouseName	studentID	firstName	lastName	gender	bloodStatus	hairColor	eyeColor
Filius Flitwick	Ravenclaw	20	Cho	Chang	Female	Pure-blood	Black	Dark

Business Question 15: The school is looking if there are boys who are pure blooded and have hair color as red, so that they can be cast as Santa Claus in the school's annual 'Muggle Christmas' Play.



number of boys
3

Business Question 16: Which club has the most number of students? The School Event Organization team is looking for volunteers to help organize the event and they are looking for the clubs who have the most number of members with them.



Result Grid | Filter Rows: | Exp

ClubID	ClubName	MAX(`Number of students`)
28	Chess Club	5

Business Question 17: How many 3 credit classes are there in the 7th year? A teacher is looking

for the number of classes with 3 credits in the semester of 7th year class in order to make changes to the next semester.

	Number of 3 credit classes
▶	4

Business Question 18: List of 2nd year students got an 'O' grade? Faculty is interested in knowing the students who got O grades in the 2nd year class in order to celebrate their success by keeping them in the elite club of the college.

	studentName	levelName	GradeLetter
▶	Oliver Wood	Second Year	O
	Michael Corner	Second Year	O
	Vincent Crabbe	Second Year	O
	Scorpius Hyperion Malfoy	Second Year	O

Business Question 19: Which professor teaches the most number of students? School administration is looking to reduce the stress of the professors who have the highest number of students teaching in order to maintain work life balance.

	employeeFirstName	employeeLastName	Number of students
▶	Filius	Flitwick	52
	Rubeus	Hagrid	52
	Minerva	McGonagall	52
	Mulch	Higgins	42
	Aurora	Sinistra	40
	Bathsheda	Babbling	40

Result 16 ×

Business Question 20: How many pets are above the average pet cost? School is looking to buy more pets for the school and to decide which pets to buy they want to know how many pets are above the average pet cost.

	Number of pets
▶	5

Business Question 21: A iCharm company is visiting campus to place the students who have learn most charms. They want to build a Select statement to find which course teaches most number of charms. Please help iCharm write the SQL statement: -- Which class teaches the most spells

className	NUMBER_OF_Spells
Defence Against the Dark Arts IV	34
Charms III	32
Charms I	29
Charms IV	27
Defence Against the Dark Arts V	23
Charms V	22
Defence Against the Dark Arts III	21
Charms VI	18
Charms II	16
Charms VII	14

Business Question 22: iPetStore is planning to open a new pet store in town and they are doing research on what the most owned pet in Hogwarts is. Can you please help iPetStore build an inventory of most owned pet type:

animalGroup	NumberOfPets
Bird	5

Business Question 23: Parvati got expelled from the hostel because she was unethically using the charms she learnt in class in non-magic spaces. As she is now only attending Hogwarts on Tuesday, she is looking for a 3-credit course to register that is offered only on Tuesday.

classID	className	subjectID	classLevelID	professorID	credits	location	scheduledDay	classTime
TNFG-101	Transfiguration I	TNFG	1	22	3	Classroom 34	Tuesday	16:00:00
TNFG-501	Transfiguration V	TNFG	5	22	3	Classroom 1B	Tuesday	14:00:00
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Business Question 24: Voldemort has cast a spell on all students who are “Half-Blood” and have blue eyes. Can you please create a list of such students so that they can be put under special supervision.

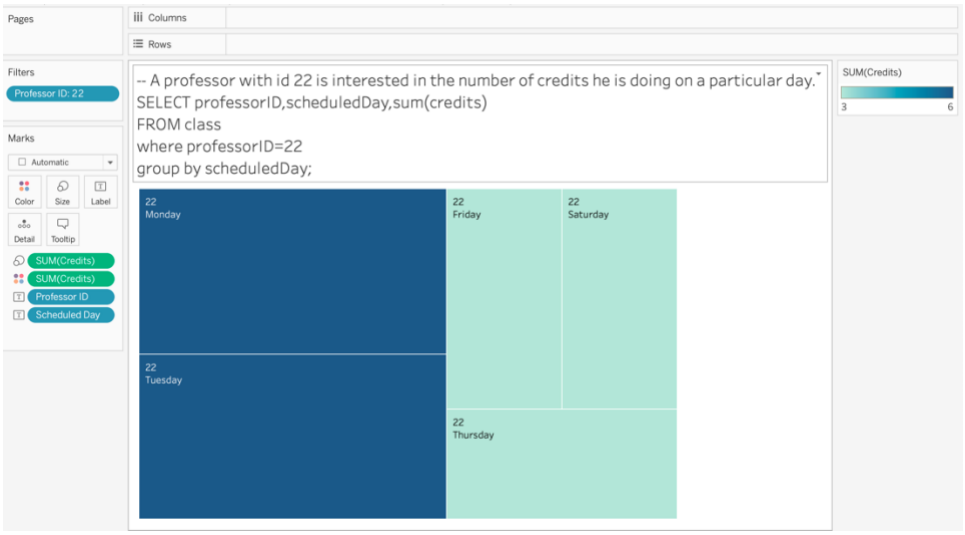
	firstName	lastName
▶	Rose	Granger-Weasley

Business Question 25: Hogwarts is planning to recruit new faculty because of a surge in new admissions, they want to check the diversity ratio for female to male for all professors. If the diversity ratio is >2 then recommend hiring female or else hire Any.

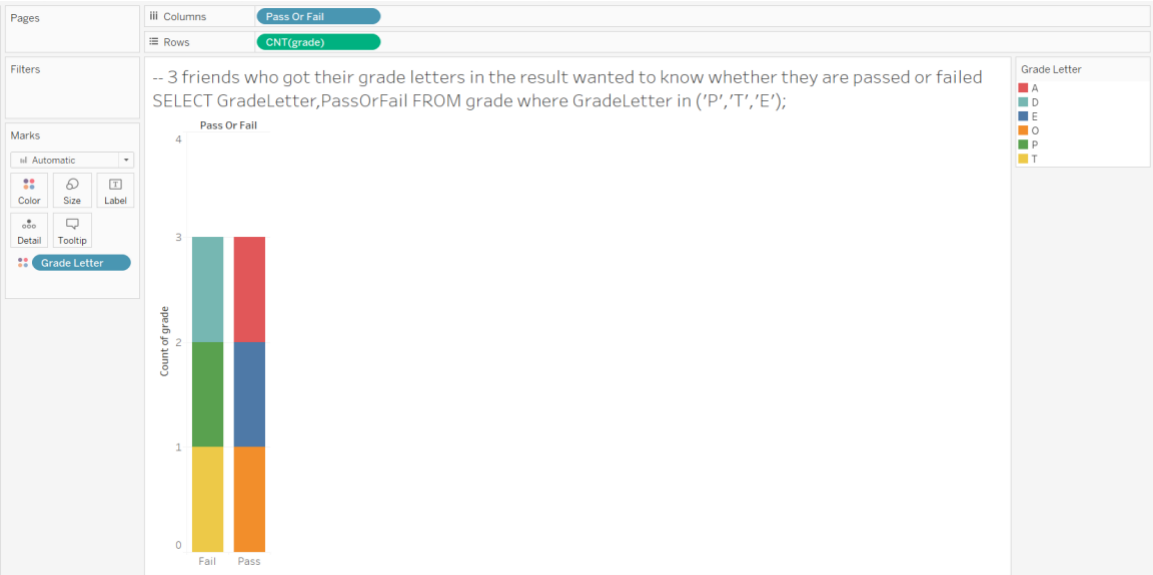
	MALES	FEMALES	Diversity_ratio	IF ((COUNT(CASE WHEN gender='MALE' TH...
▶	22	10	2.2000	HIRE FEMALE

Data Visualizations Scenarios, with corresponding screenshots:

Visualization 1: Professor Minerva McGonagall with id 22 is interested in the number of credits she is doing over a particular week. To answer this business question 11, we used Tableau to create a heat map, using `ProfessorID` and `ScheduleDay` as the dimensions and `Credits` as the measure.

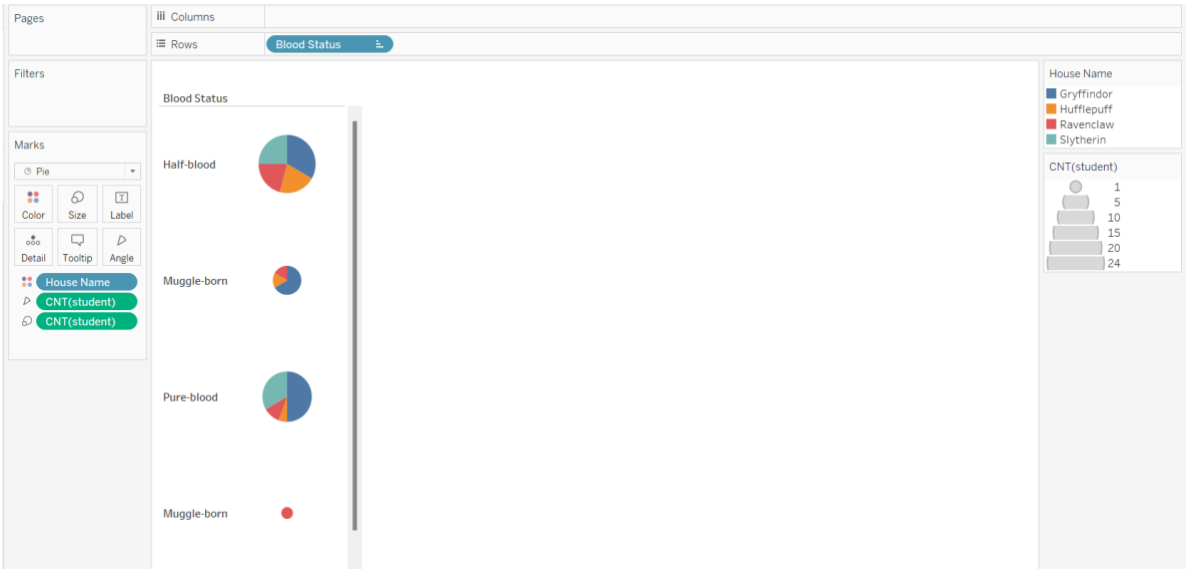


Visualization 2: 3 friends who got their grade letters in the result wanted to know whether they have passed or failed so that they can apply for their internships at Ministry of Magic and calculate the placement rate of Hogwarts. To answer this business question 12, we used Tableau to create a stacked bar graph, using `gradeLetter` and `PassorFail` as the dimensions, to get the total of the grade.



Visualization 3: Administration is interested in knowing the number of classes taken by each professor in this semester. Business Question 13: The administration is interested in knowing the number of classes taken by each professor this semester. Here, we joined the `Employee` table to the `Classes` table, to get the dimensions of `employeeType`, `employeeName` and then get the customized count of classes taught by each professor.

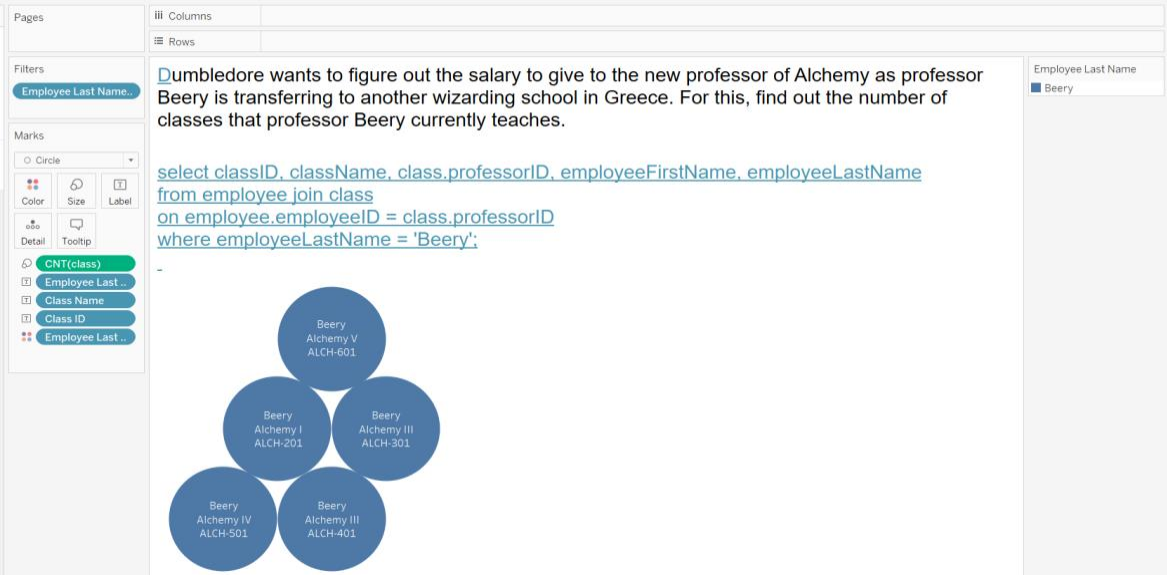
Visualization 4: Business Question 14: Administration wants to find the distribution of blood groups of students in each house. We used the `Students' table, with `bloodstatus` taken as the dimension and grouped by 'houseID' to find the pie charts. To label the charts, the `house` table was joined with the `Students` table and displayed in th legend.



Visualization 5: In response to Business Question 15, where the school is looking if there are boys who are pure blooded and have hair color as red, so that they can be cast as Santa Claus in

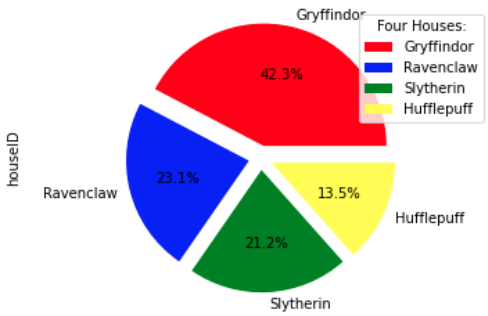
the school's annual 'Muggle Christmas' Play, we have used a bar charts to quantify each student's name

Visualization 6: Dumbledore wants to figure out the salary to give to the new professor of Alchemy as Professor Beery is transferring to another wizarding school in Greece. For this, find out the number of classes that Professor Beery currently teaches.

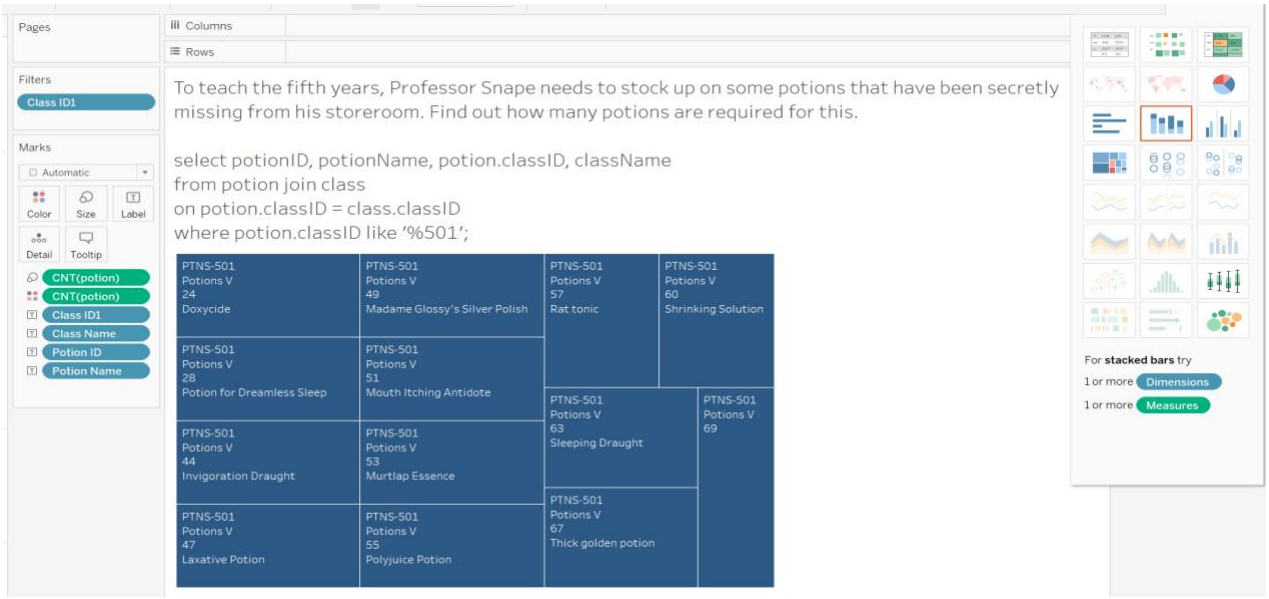


Visualization 7: The percentage of students in each house visualized in python will help the sorting hat to determine how best to place the students in the upcoming years. We joined the `house` table to the `student` table and used the pie chart to represent the percentage of students.

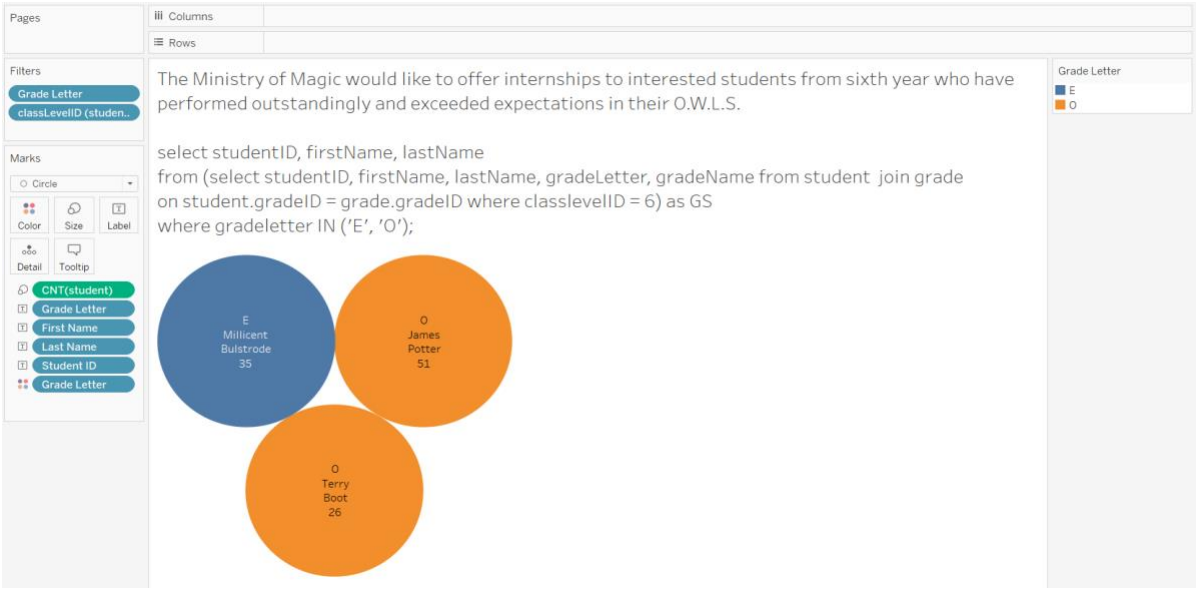
```
Out[13]: <matplotlib.legend.Legend at 0x119356550>
```



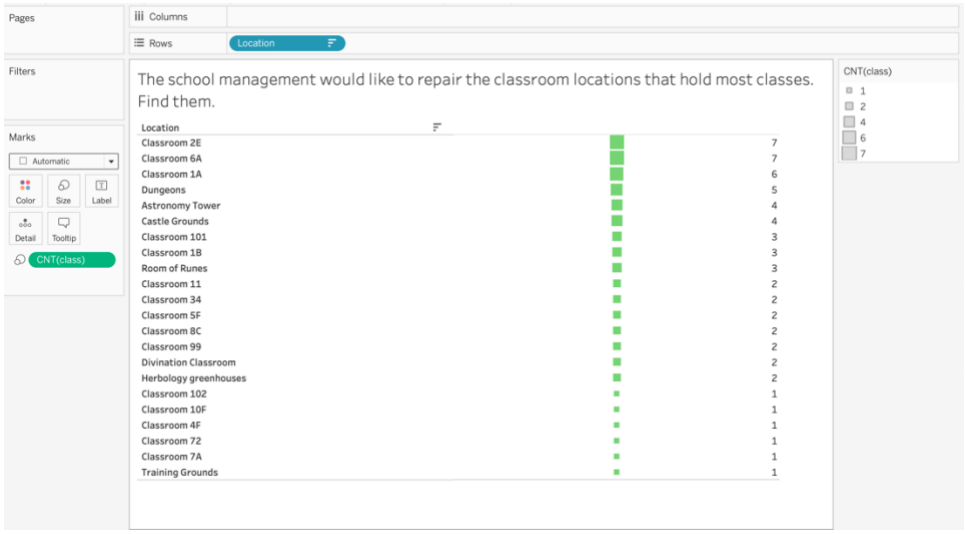
Visualization 8: To teach the fifth years, Professor Snape needs to stock up on some potions that have been secretly missing from his storeroom. Find out how many potions are required for this. A Square map was used to show this



Visualization 9: The Ministry of Magic would like to offer internships to interested students from sixth year who have performed outstandingly and exceeded expectations in their O.W.L.S.

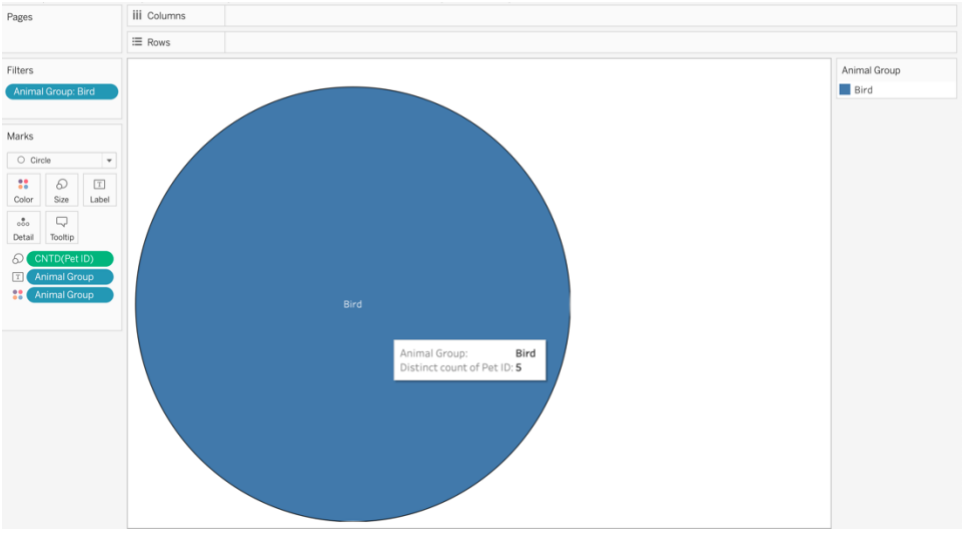


Visualization 10: The school management would like to repair the classroom locations that hold most classes. Find them.

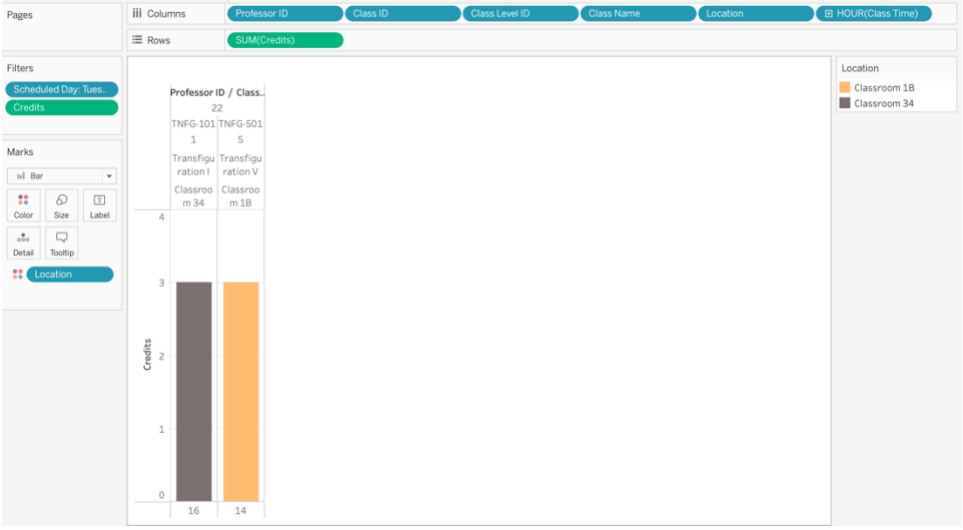


Visualization 11: A iCharm company is visiting campus to place the students who have learn most charms. Find which course teaches most number of charms. A highlighted table was used to show the scale.

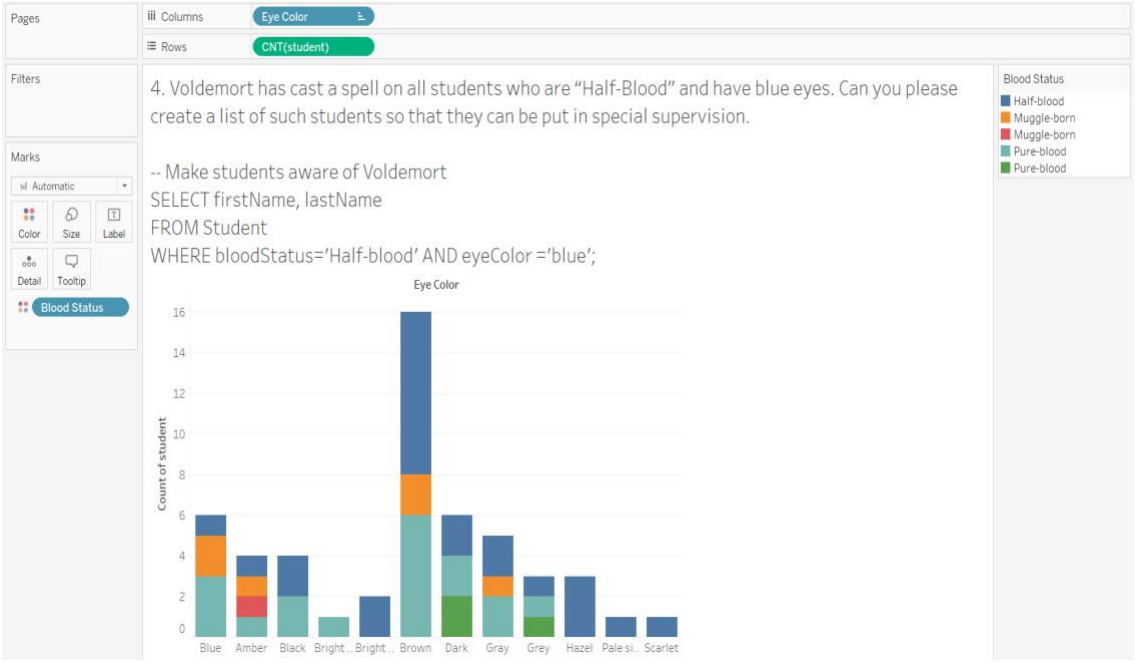
Visualization 12: iPetStore is planning to open a new pet store in town and they are doing research on what is the most owned pet in Hogwarts. Can you please help iPetStore build an inventory of most liked pets:



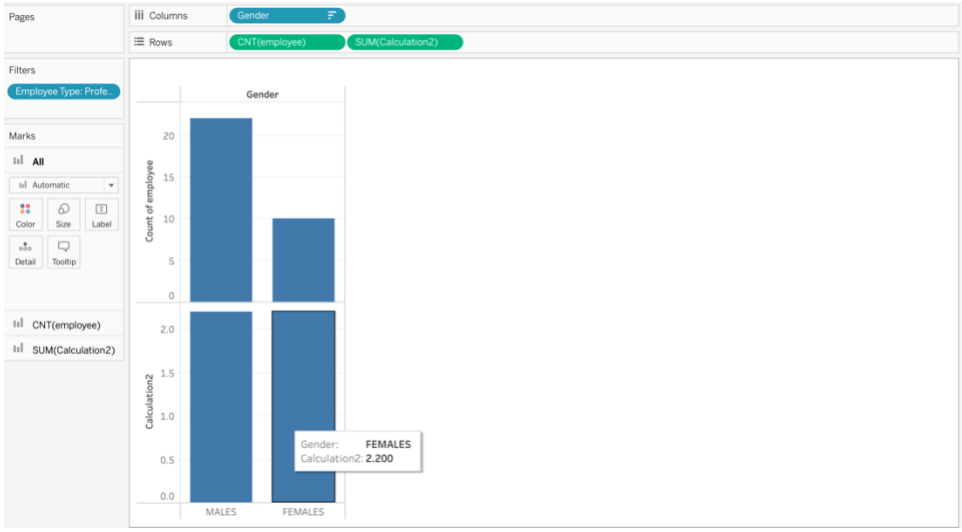
Visualization 13: Parvati got expelled from the hostel because she was unethical using the charms she learnt in class. She is now only visiting Hogwarts on Tuesday, she is looking for a 3 credit course to register that is offered only on tuesday.



Visualization 14: Voldemort has cast a spell on all students who are "Half-Blood" and have blue eyes. Can you please create a count of students so that they can be put in special supervision.

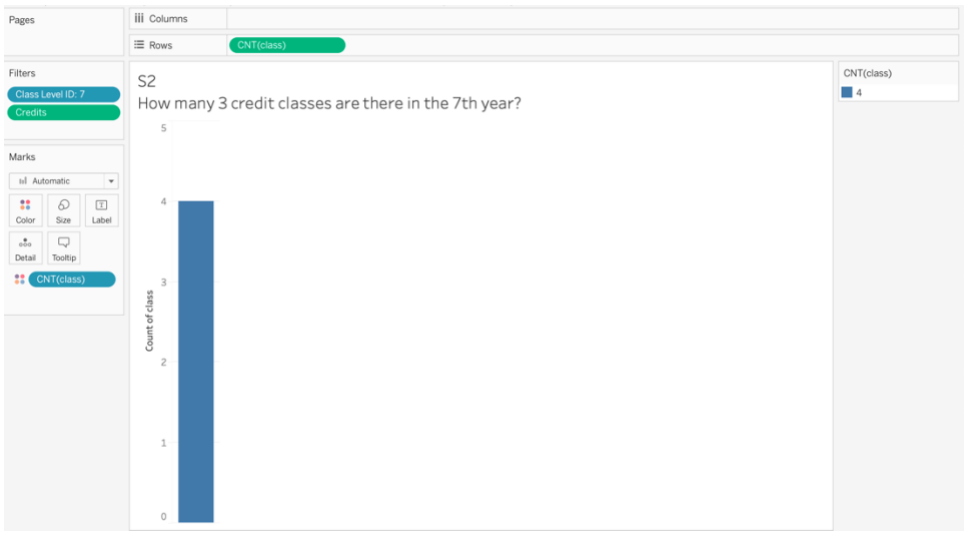


Visualization 15: Hogwarts is planning to recruit new faculty because of a surge in new admissions, they want to check the diversity ratio for female to male for all professors. If the diversity ratio is >2 then recommend to hire female or else hire Any.

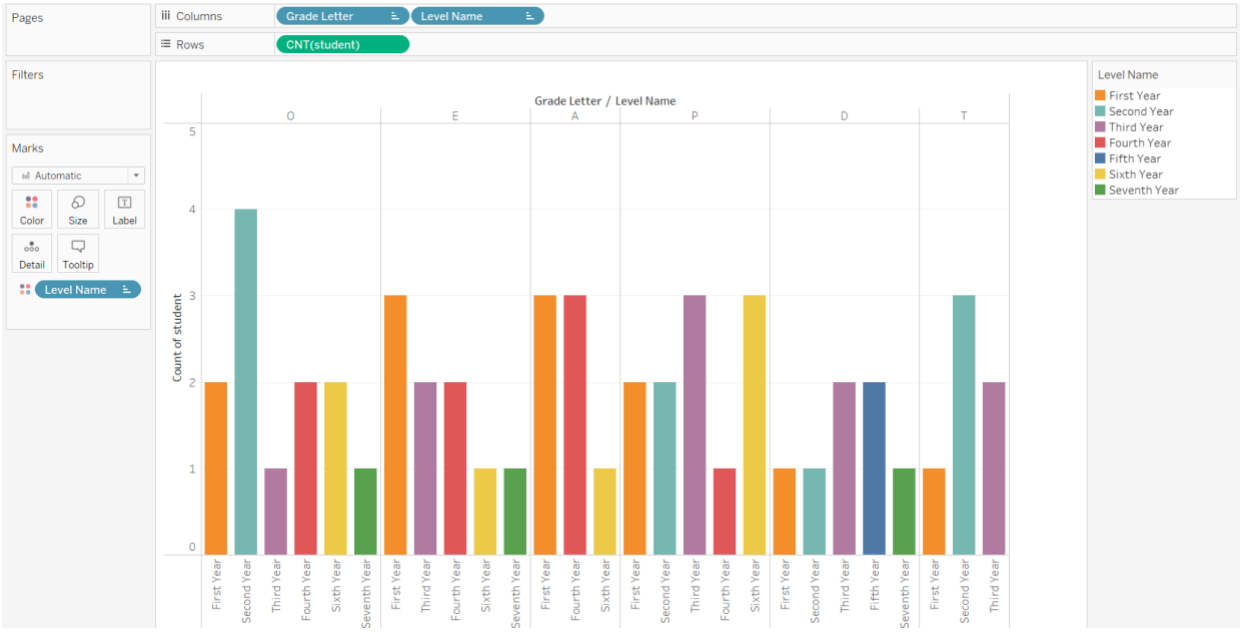


Visualization 16: The School Event Organization team is looking for volunteers to help organize the event and they are looking for the clubs which have the most number of members with them.

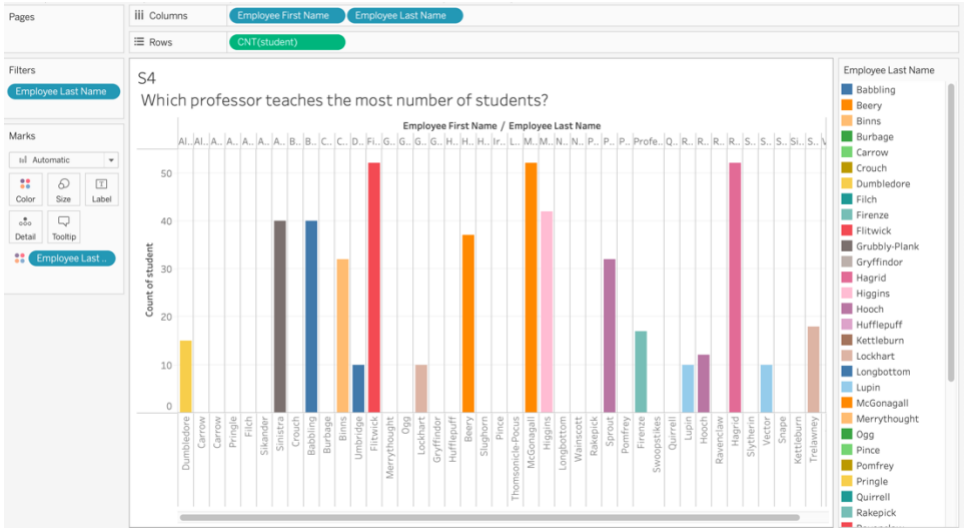
Visualization 17: A teacher is looking for the number of classes with 3 credits in the semester of 7th year class in order to make changes to the next semester.



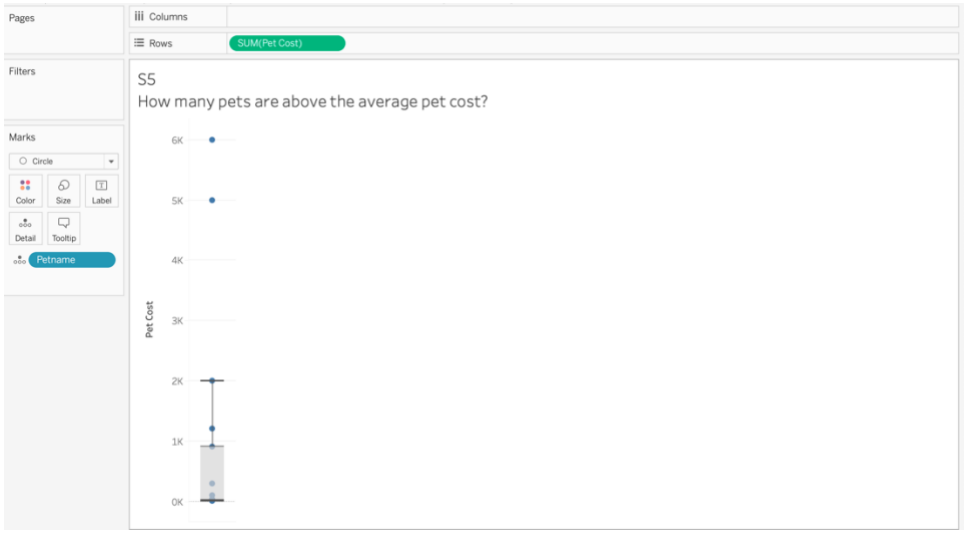
Visualization 18: The faculty wants to check the distribution of grades of the students in order to curve the grades for publishing the final results.



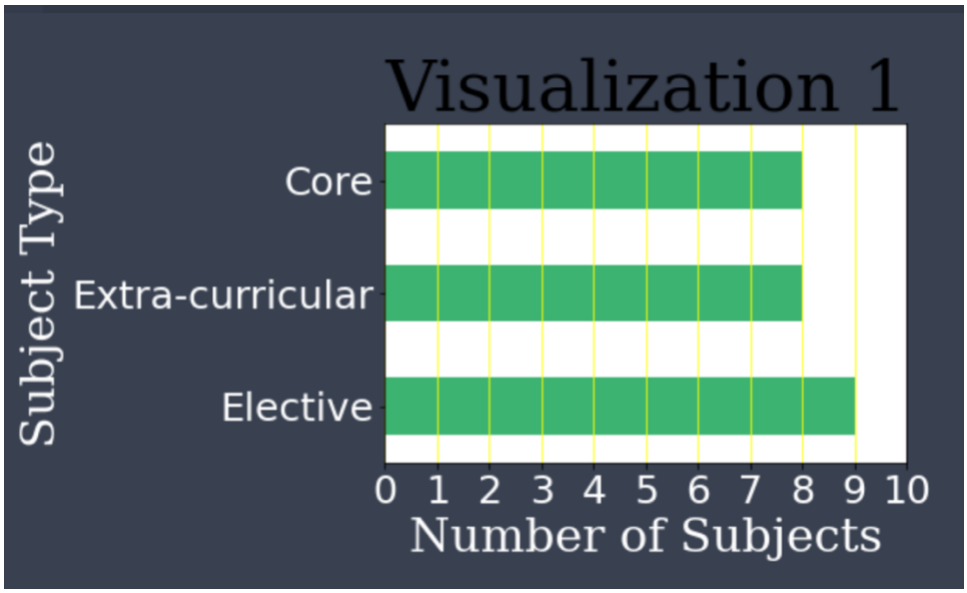
Visualization 19: School administration is looking to reduce the stress of the professors who have the highest number of students teaching in order to maintain work life balance.



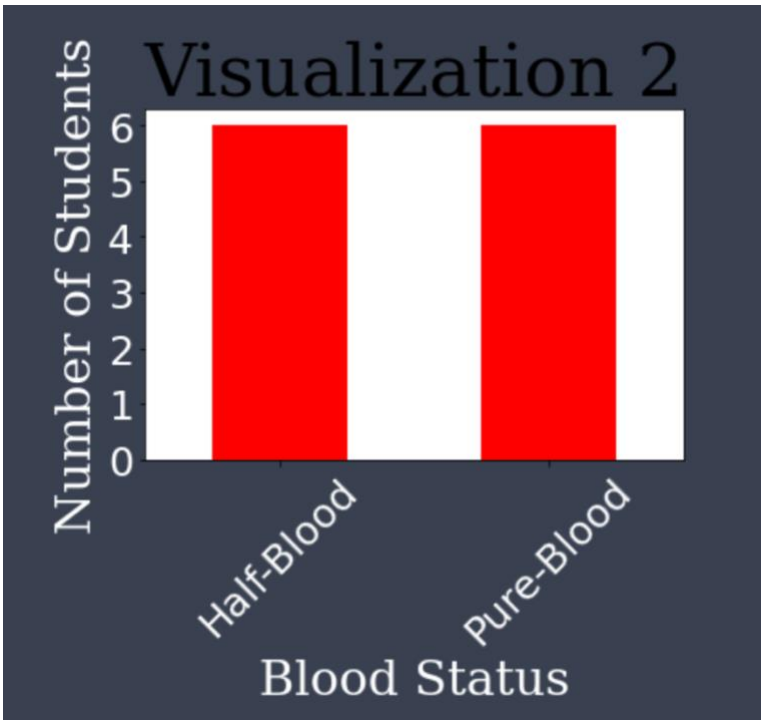
Visualization 20: School is looking to buy more pets for the school and to decide which pets to buy as they want to know how many pets are above the average pet cost.



Visualization 21: The headmaster thinks that they need to revise the school syllabus and wants to check how many subject types are currently available at Hogwarts. To depict this, we have used a Horizontal Bar Plot in Python, with the subject type on the left and the count of each subject type on the bottom. The dimensions used are the `subjectType` column from the `Subjects` table and its aggregated count.

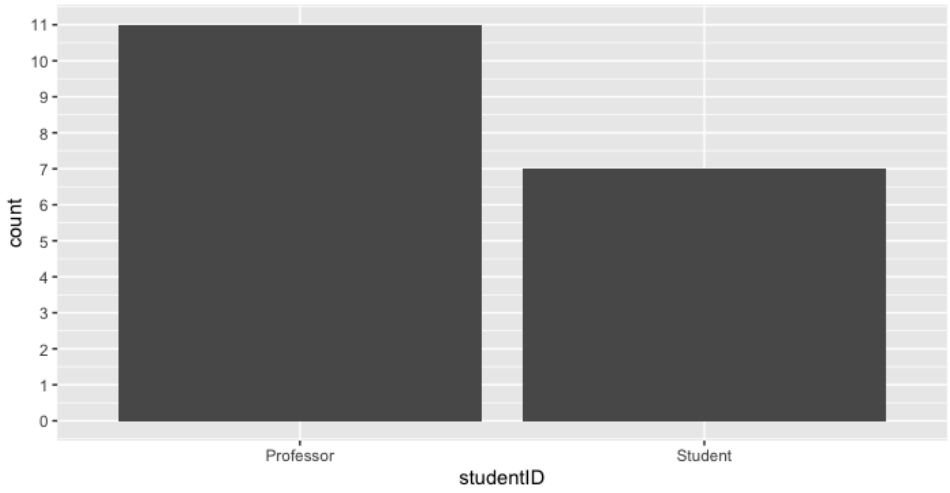


Visualization 22: The head of Slytherin wants to find all the Blood Status of his member students to put in his final house diversity report. To achieve this, we joined the `house` table to the `student` table, filtered by House Name (Slytherin) and used the plot function in Python to create a Vertical Bar Plot using the dimensions of Blood Status and its count.

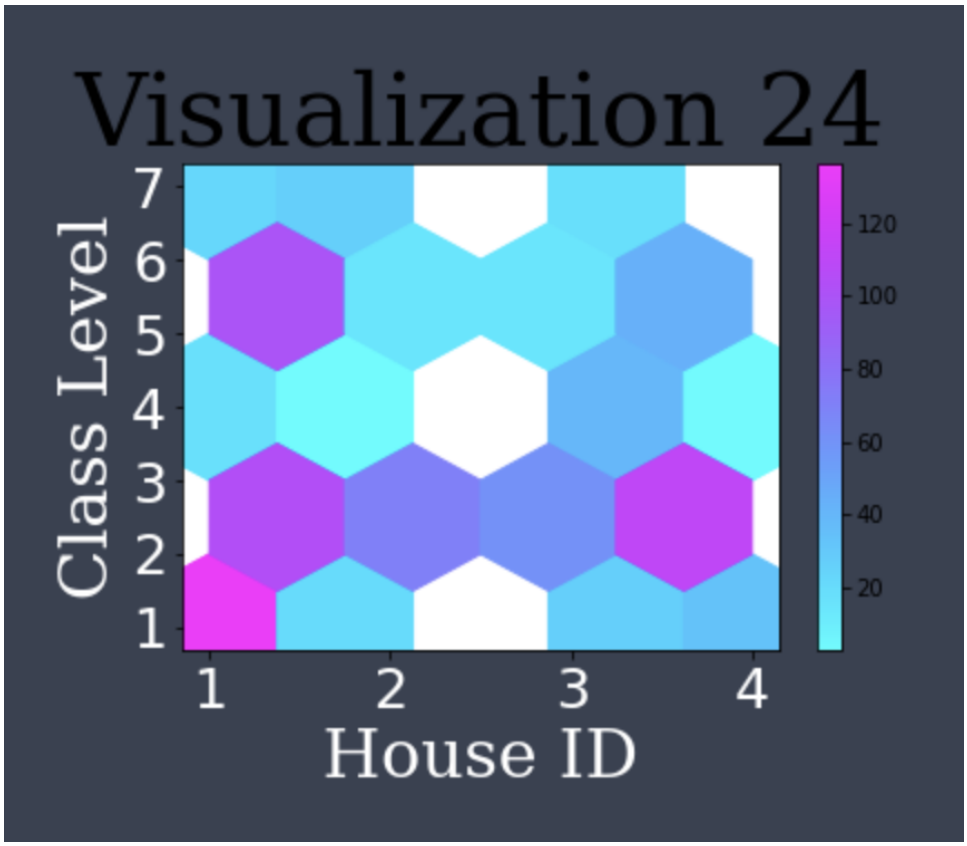


Visualization 23: The number of pets owned by professors vs students visualized in R studio, using the ggplot library. This bar plot shows the type of owner in the x-axis and the group

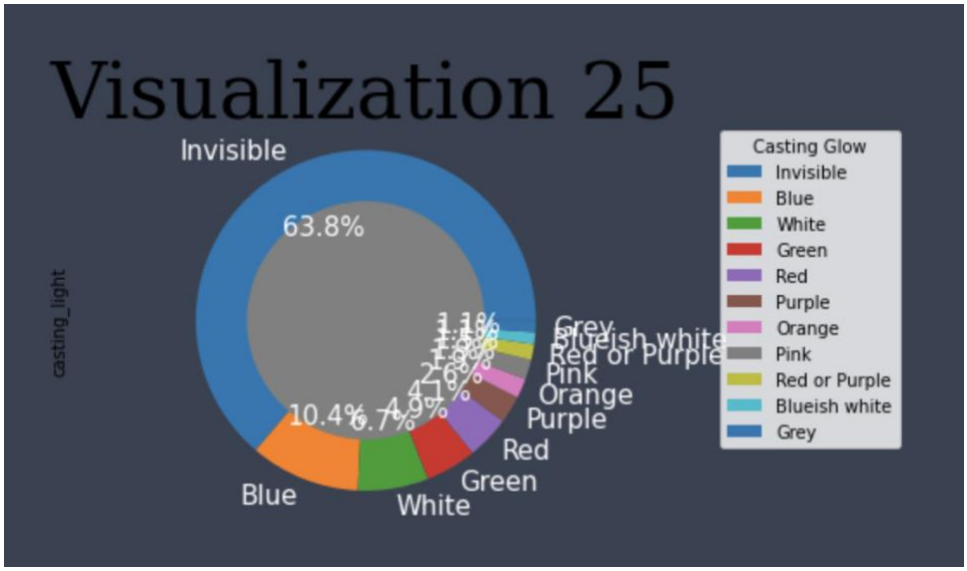
aggregate along the y-axis. Having this visualization answers Business Question 3, where the school decides whether to build another building to house pets owned by Student.



Visualization 24: This hexbin graph shows the total number of club members across all Houses and Class levels. The seven class levels have been plotted along the y-axis and the 4 Houses have been plotted along the x-axis. The legend on the left shows the members of the club, denoted by club ID.



Visualization 25: This visualization shows the percentiles of the casting light of various spells from the `Spells` table. In order to maintain visibility, spells which have unique casting lights have been filtered out. This donut chart was made using python using the `castingLight` attribute from the `Spells` table.



Bibliography And Team Contributions:

Data Sources

We mostly sourced our data from the Harry Potter Dataset found on Kaggle
https://www.kaggle.com/gulsahdemiryurek/harry-potter-dataset

We also manually enriched our datasets using information from the Harry Potter Wiki along with some assumptions of our own (with the help of the Hogwarts staff).

Team Member Contributions

Number	Name	SQL Contributions	Visualization Contribution
1	Adrija Barvadheesh	Question 1 Question 2 Question 3 Question 4 Question 5	Visualization 21 Visualization 22 Visualization 23 Visualization 24 Visualization 25
2	Karteeka Uppalapati	Question 6 Question 7 Question 8 Question 9 Question 10	Visualization 6 Visualization 7 Visualization 8 Visualization 9 Visualization 10
3	Jay Annangi	Question 11 Question 12 Question 13 Question 14 Question 15	Visualization 1 Visualization 2 Visualization 3 Visualization 4 Visualization 5
4	Sreesh Sattiyamourthy	Question 16 Question 17 Question 18 Question 19 Question 20	Visualization 16 Visualization 17 Visualization 18 Visualization 19 Visualization 20
5	Harvey Singh	Question 21 Question 22 Question 23 Question 24 Question 25	Visualization 11 Visualization 12 Visualization 13 Visualization 14 Visualization 15