



Sunshine Valley

MAGICAL PONY FARM

MALVIKA BODH

ebay Analyst Position Interview Task | 07/04/2019

Introduction



Image Credit: AIHOME

Sunshine Valley is a struggling mid-level commercial pony farm, located in the far away land of Ponia. Home to 15,000 ponies, each identifiable by a unique numerical ID, the farm has been recently purchased by our company, and we have acquired the farm's data. The data includes 4 tables, which respectively describe individual pony characteristics (such as name, date of birth, breed, etc.), the associated magical powers (0 to 4), the relationships between ponies (friends, enemies, frenemies, and family), and their treatment histories (physical, psychological or none). Each of these parameters may influence the market worth of individual ponies.

Here, I provide a broad overview of the role of attributes in affecting the worth of a pony:

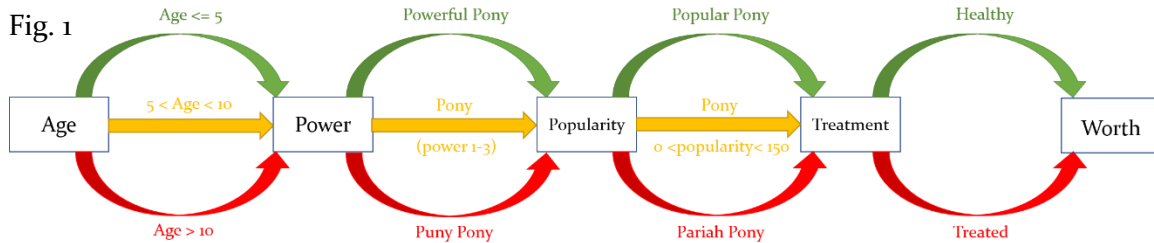
- **Age** (worth peaks around age of 5, then rapidly declines; little worth if age > 10)
- **Magical Powers** (more powers a pony has, more its worth)
 - Powerful Pony (all 4 powers): high worth
 - Puny Pony (1 power at age > 5): little worth
- **Popularity** (greater the popularity, more its worth);
 - Depends on relationship with other ponies, type of magical powers, and number of family members on farm
 - Ponies with popularity ≥ 150 considered Popular (high worth)
 - If popularity is negative then considered Pariah (no worth, and degrades market worth of other ponies)
- **Treatment History** (more treatments lead to more costs, hence lower worth)
 - But, if pony has Power of Gratitude, then no costs – hence, no effect on worth

The task is to use the dataset to provide preliminary guidelines for improving the situation on farm, thus maximizing profit from sales. For this report, given the volatility of the pony market, I will focus on providing a rough estimate of the worth of Sunshine Farm, as well as establishing the guidelines for a more robust estimation in future (based on accurate financial details).

Take Stock

In this analysis, I am interested in determining the market worth of Sunshine Farm based on the attributes of **Age Groups**, **Total Powers**, **Popularity**, and **Treatments**.

Fig. 1



Given the lack of actual costs and revenues (or value) for each of these aspects of pony farming, I will make specific assumptions based on the provided language in the project sheet. Notably, the calculation of market worth is broken down into the 4 categories mentioned above (See Fig 1: **Green**: Positive effect; **Red**: Negative effect), starting with age.

1. Age

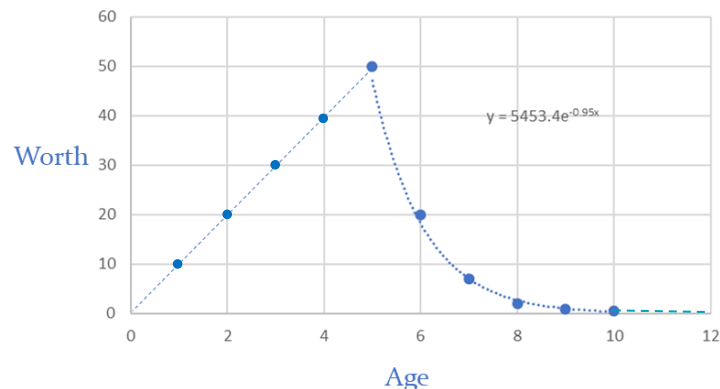
Ponies have a life expectancy of 25 years, with worth increasing till age 5, then decreasing rapidly to low worth at age 10 and beyond. For this parameter, I have categorized the data into separate bins (bin size of 1 till age 8, 2 for age 8-10 & ponies > 10 considered together), as illustrated in Fig. 3a below. A large number of ponies lie between the ages of 2-4 (**6,345**), followed by **1,410** ponies close to the peak worth age group of 4-5. Beyond the age of 10, the total number of ponies falls to **749** (detailed statistics available in SQL code).

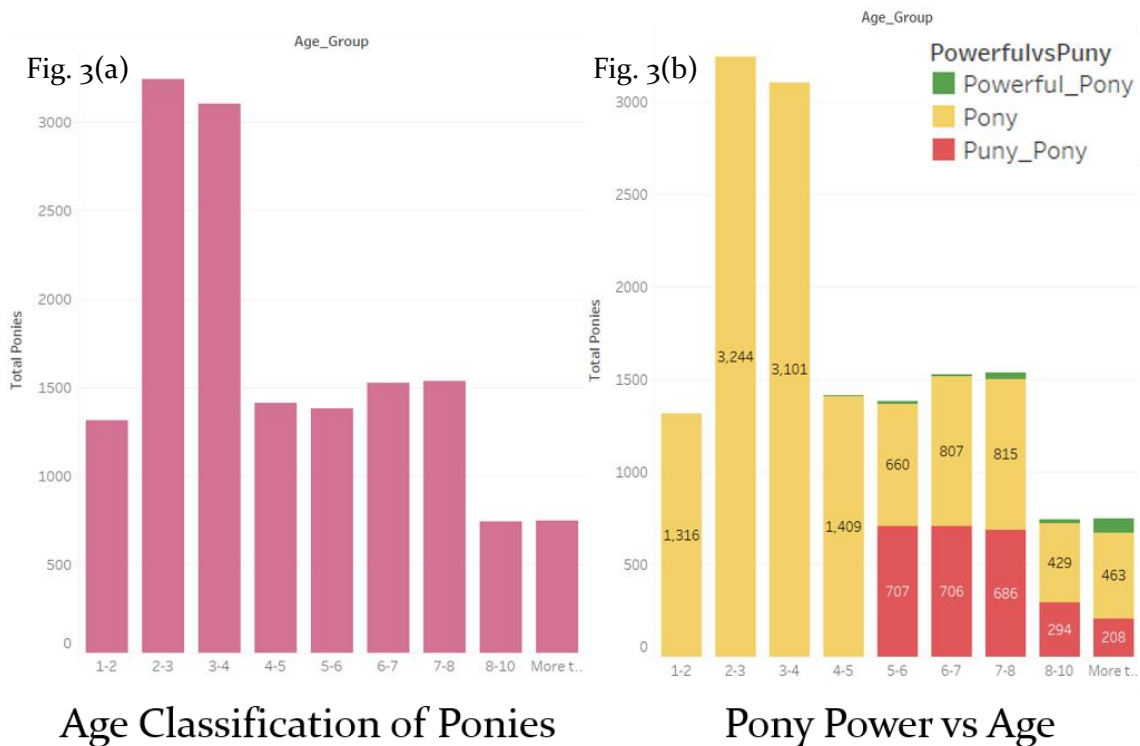
To estimate the **market worth** (in Ponia's currency of silver pieces) **solely due to value addition by age**, I make the following assumptions:

- Initial value of pony is zero at birth (age 0-1)
- For ages 1-5, **linear** increase in value by 10 for each bin
 - Thus, a pony at age 5 (peak) has a value of 40
- Beyond age of 5, there is a rapid **exponential** decay of value (Ae^{-x} : See Fig. 2)
 - At age 6, pony has a value of 20
 - At age 10 and beyond, value close to zero ('little' worth)

Assumed behavior of Worth vs Age

Fig. 2





Using these assumptions, and the values acquired from final SQL table (displayed in Fig. 3a), I calculate initial worth due to age as:

$$\text{Worth_Age} = \text{SUM (No. of Ponies per age category * Assigned value of each category)}$$

Thus, I get $\text{Worth_Age} = (1316 * 10 + 3244 * 20 + 3101 * 30 + 1410 * 40 + 1380 * 20 + 1524 * 7 + 1,536 * 2 + 740 * 1 + 749 * 0.5) /*Refer to Analysis Sheet*/$

Or, **Worth_Age = 269,925**

It is interesting to note that the age group **3-4 has the maximum worth (93,030)** due a higher number of ponies, with second highest value of age category. However, this is the first step of the calculation of farm worth.

2. Power

For the second part, I will consider the value addition of power to the market worth, in addition to Worth_Age which was already computed. The categorization of ponies here is based according to number of powers and age (illustrated in Fig. 3b, and Fig. 4).

The data displays an (expected) increase in number of powers with age, with most powerful ponies (no. of powers = 4) found in the age category of more than 10 (**78 ponies**). For number of powers = 1, we see a peak in pony count for age groups 2-4 (1458 and 1487 for the two groups respectively), which falls to 573 in the peak age range of 4-5. This

indicates that a substantial number of ponies acquire at least one power at the age of 2, and some additional powers with increase in age. This requires future investigation to determine at what age most ponies transition from one power to multi power ponies.

Age Group vs Total Power & Popularity

Fig. 4

Age_Group	PonyPopularity	Total_Powers					Grand Total
		0	1	2	3	4	
1-2	Null	15					15
	Pariah Pony	222					222
	Pony	889					889
	Popular Pony	190					190
2-3	Null	8	35	5			48
	Pariah Pony	92	231	23			346
	Pony	452	1,458	242			2,152
	Popular Pony	105	461	132			698
3-4	Null	2	24	8	1		35
	Pariah Pony	32	245	45			322
	Pony	169	1,487	360	29		2,045
	Popular Pony	42	479	164	14		699
4-5	Null	2	9	3			14
	Pariah Pony	14	98	28	1		141
	Pony	82	573	245	33	1	934
	Popular Pony	23	182	94	22		321
5-6	Null		8	6	1		15
	Pariah Pony	20	84	33	7		144
	Pony	77	475	290	55	5	902
	Popular Pony	16	140	123	32	8	319
6-7	Null	1	8	9	3		21
	Pariah Pony	18	69	46	3		136
	Pony	88	456	312	86	6	948
	Popular Pony	22	173	168	51	5	419
7-8	Null	2	2	5	3		12
	Pariah Pony	23	72	32	8		135
	Pony	72	466	317	108	19	982
	Popular Pony	20	146	148	77	16	407
8-10	Null		4				4
	Pariah Pony	8	34	15	5		62
	Pony	32	186	191	50	8	467
	Popular Pony	9	70	80	39	9	207
More than 10	Null		2		1	2	5
	Pariah Pony	11	27	13	3		54
	Pony	58	130	159	96	46	489
	Popular Pony	4	49	64	54	30	201
Grand Total		2,820	7,883	3,360	782	155	15,000

To estimate the **market worth solely due to the number of powers**, I need to incorporate the increase in worth with each power for ponies in each age category. This is in addition to Worth_Age.

I will make the following assumptions here:

- Increase in value by **10** for each power
- Missing power values are considered as zero

- Thus, a powerful pony (155 total) will have an additional value of +40, apart from the value due to its age
- A puny pony (2601 total) will have an additional value of 10, which when considering the value due to age leads to a relatively low worth (consistent with program sheet)

Again, using these assumptions, and the values acquired from final SQL table (displayed in Fig. 4, 5), I calculate initial worth due to age as:

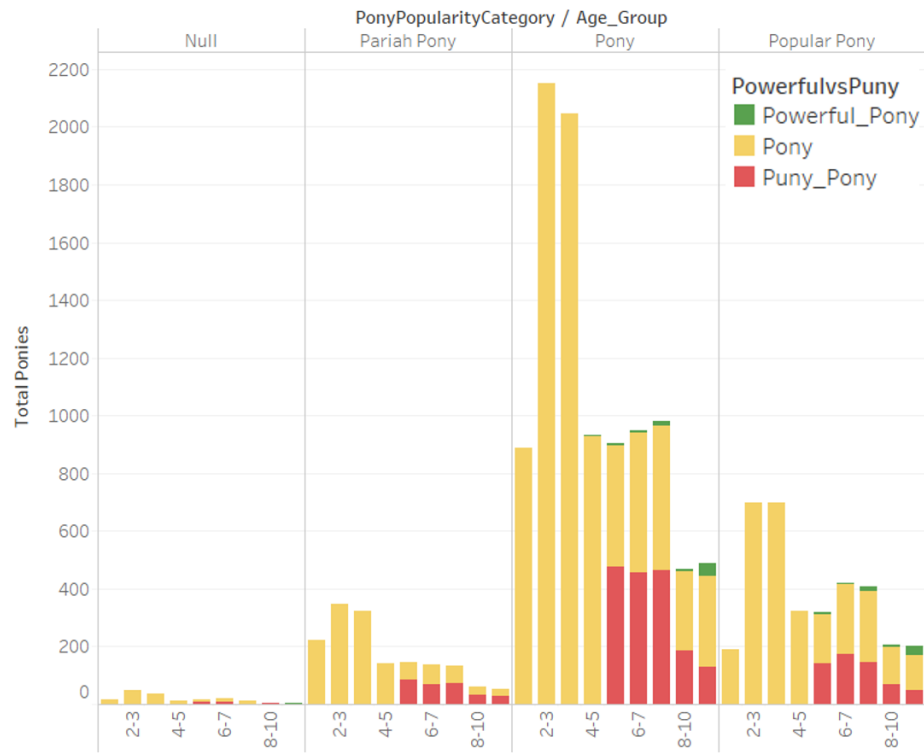
$$\text{Worth_Age_Power} = \text{Worth_Age} + \text{SUM (No. of Ponies per power * Assigned value of each power category)}$$

Thus, I get $\text{Worth_Age_Power} = 269,925 + (0 + 7883 * 10 + 3360 * 20 + 782 * 30 + 155 * 40)$

OR, $\text{Worth_Age_Power} = 445,615 (269,925 + 175,690)$

Thus, the worth addition due to power is also substantial (~40% of present estimated worth) based on my assumptions.

Fig. 5



No. of Ponies vs Age and Popularity (Pony Power in color)

3. Popularity

In the third part, I incorporate the effects of popularity on the market worth of Sunshine Valley. This will be added on to the `Worth_Age_Power` value from the previous section. As illustrated in Fig. 5 and 6b, the breakdown of total number of ponies by age, power, and popularity indicates that most ponies (9,808) have a popularity between 0 to 150. Further, most of these ponies have powers between 0-3.

There are 3,461 popular ponies (popularity > 150) and 1,562 pariah ponies (popularity < 0), both of which impact the overall worth of the farm. To estimate the **market worth solely due to popularity**, I make the following assumptions:

- Considering a min/max popularity score interval of ~1200, I assume an increase in value of **1 for each 100** points of popularity
 - Popular ponies get an additional +1 per 100 points of popularity (“worth considerably more”)
- For **negative popularity**, I assume a **decrease of 1 for each 1000 popularity points** (pariah pony) to all ponies (as it degrades value of all others). This assumption is made to ensure a reasonable decrease compared to current worth
- Finally, a pariah pony is worth nothing, so I need to subtract the value addition so far for all pariah ponies (due to age and power) (see Fig. 4)

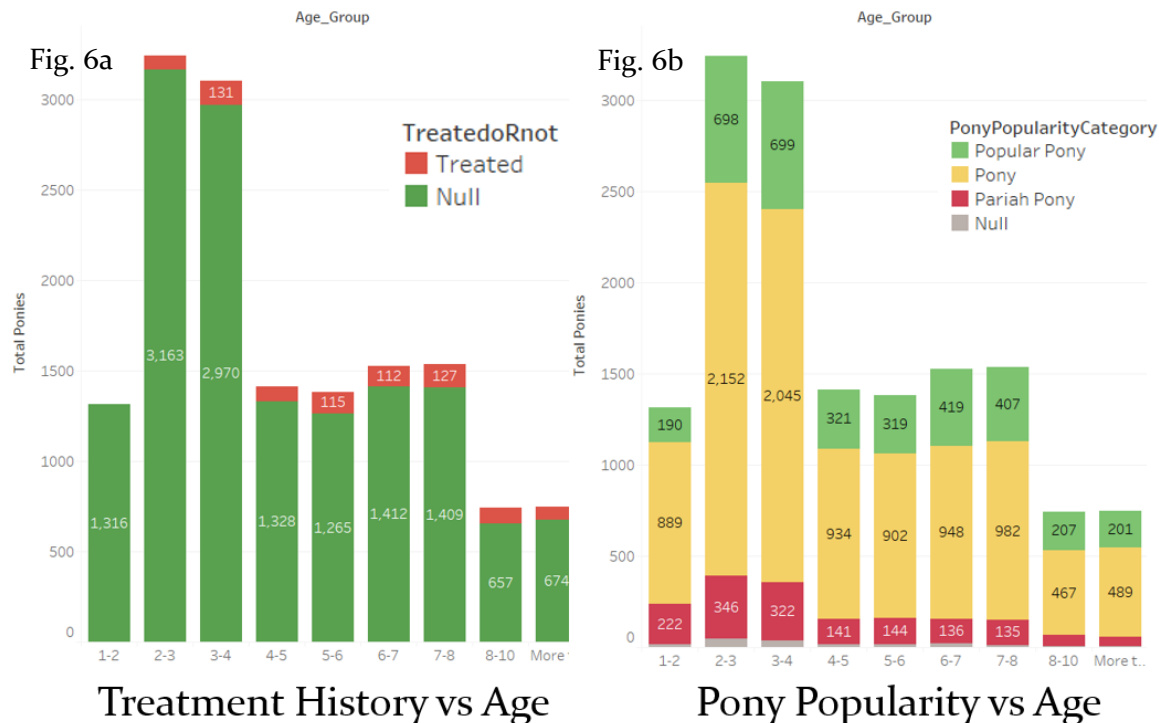
Using these assumptions, I incorporate the effect of market worth due to popularity to get overall worth as:

$$\text{Worth_Age_Power_Popularity} = \text{Worth_Age_Power} + \text{SUM (No. of popular/regular ponies * Assigned value for popularity category + No. of pariah ponies * Assigned negative value * Total ponies affected)} - \text{Worth_Age_Power (Pariah Ponies)}$$

Thus, I get $\text{Worth_Age_Power_Popularity} = 445,615 + \text{/*SUM*/} (3461 * (2/100) + 9808 * (1/100) + 1562 * (-1/1000) * (14,999)) - \text{/*Power Value of Pariah*/} (0 + 860 * 10 + 235 * 20 + 27 * 30 + 0) - \text{/*Age value of Pariah*/} (222 * 10 + 346 * 20 + 322 * 30 + 141 * 40 + 144 * 20 + 136 * 7 + 135 * 2 + 62 * 1 + 54 * 0.5)$

$$\text{OR, Worth_Age_Power_Popularity} = 379,698 (445,700 - 23,261 - 14,110 - 28,631)$$

Popularity appears to have a significant effect on the market value of the farm, due to the mutual increase in value for ponies with positive popularities.



4. Treatment History

Finally, I will incorporate the final parameter of treatment history (See Fig. 6a), with consideration for the number of treatments undergone. I have investigated this attribute with behavior with popularity, and power (see attached code). There are 806 ponies which have undergone a total of 1051 treatments at cost to the farm. Most 'Treated' ponies have a popularity between 0-150. As expected, no powerful pony is considered treated (due to The Power of Gratitude).

For this estimation, I will assume:

- a) Both physical and psychological treatments are equally detrimental to value
 - Same, significant cost of 10 (decrease in value)
- b) Multiple treatments costs the same on the same pony
- c) Treatment cost is same irrespective of age

To estimate the final market worth of the farm, we subtract the costs of treatment from `Worth_Age_Power_Popularity`, using the simple equation:

$$\text{Market_Worth} = \text{Worth_Age_Power_Popularity} - \text{SUM (No. of treatments * Cost of a treatment)}$$

Thus, I get `Market_Worth = 379,698 - (1051 * 10)`

OR, `Market_Worth = 369,188`

The estimated market worth of Sunshine Valley is ~370,000 silver pieces

Find Anomalies

The combined farm dataset, and subsequent analysis reveals a few anomalies, either in the form of absent data or suspicious values which require a second check:

- a) Power Data:
 - i. Missing for certain ponies, which makes it hard to consider worth/cost
 - Do they have no powers or is the data missing?
 - ii. Date when powers were acquired is not provided – this is important to calculate accurate values for popularity of ponies
 - E.g. A pony may have already been treated at cost to the farm and may later acquire the power of gratitude. In this case, the model fails to include this initial cost
- b) Relationships:
 - i. No relationships mentioned for some ponies
 - Are they lone rangers or is the data missing?
 - ii. Zero values for relationship intensity for some ponies
 - Friend/ Enemy relationship type with an intensity value of zero doesn't make sense
- c) Family members for Last Name 'NastyBunch': 705
 - Is data real, as it impacts the popularity of overall family?
- d) Food:
 - i. 'Black Mushrooms' have no treated cases and no power types for ponies:
 - Is this data real or not (Great health benefits with no Power)?

Provide Actionable Insight

1. *Pariah Ponies* – I suggest removal from the farm

The data indicates 1,562 pariah ponies (popularity < 0) on Sunshine Farm. These pariah ponies have no worth of their own, and decrease the worth of other ponies on the farm (by extension, decreasing market worth of farm).

- I suggest removing all pariah ponies from Sunshine Valley
- Alternatively, future investigation should focus on how pariah ponies affect the popularity of other ponies on the farm (due to relationship type or family)
 - If total increase in pony worth is greater than decrease due to pariah pony, they can stay; otherwise, they leave

2. *Puny Ponies* – Removal based on cost of upkeep vs worth and popularity

Fig. 7

Popularity vs Power

PowerfulvsPu...	PonyPopularityCategory				Grand Total
	Popular Pony	Pony	Pariah Pony	Null	
Powerful_Pony	68	85		2	155
Pony	2,815	8,010	1,276	143	12,244
Puny_Pony	578	1,713	286	24	2,601
Grand Total	3,461	9,808	1,562	169	15,000

Number of Records
2 8,010

Similar to the insight regarding Pariah Ponies, the company needs to consider the

worth addition of puny ponies, including what they contribute to other ponies (with popularity > 0) due to 'Friends', 'Frenemies', or 'Family' relationships (Fig. 7):

- I suggest removing Puny ponies with Popularity of 'Null' and 'Pariah' (310)
- 578 Puny but popular ponies should be preserved
- The 1,713 Puny ponies with popularity between 0-150 ('Pony') can be considered on case-by-case basis of increase in total worth vs cost

3. *Tradeoff between Age & Power* - Powerful Ponies at 10 vs Peak Worth at 5

Fig. 8

Age Group vs Total Power

Age_Group	Total_Pow 1_					Grand ..
	0	1	2	3	4	
1-2	1,316					1,316
2-3	657	2,185	402			3,244
3-4	245	2,235	577	44		3,101
4-5	121	862	370	56	1	1,410
5-6	113	707	452	95	13	1,380
6-7	129	706	535	143	11	1,524
7-8	117	686	502	196	35	1,536
8-10	49	294	286	94	17	740
More than 10	73	208	236	154	78	749
Grand Total	2,820	7,883	3,360	782	155	15,000

Total No. of Ponies
2,235

As mentioned in Section 2. Power of 'TAKE STOCK', there is an observed increase in number of powers with age. This in turn increases the value of the older (age > 10), more powerful ponies as compared to younger (Age < 4 or 5), less powerful ponies. Any consideration for removal of ponies will depend on value addition to market worth.

3.1 Example case for Older Ponies (Age > 10) – depends on power and popularity

The 749 older ponies on the farm are stated to have very little worth due to their age. However, Fig 9 indicates 78 powerful ponies in this age category (with 463 regular ponies of power 0-3 (except puny), and 208 puny ponies). Further, Fig. 9 indicates 201 popular ponies (and 54 pariah ponies).

Based on these values, I suggest:

- The 30 Powerful and popular ponies increase the worth beyond the value for their age, and should be preserved
- All 78 powerful old ponies, and all 201 popular old ponies (Total: 249) have an increased worth, and should be preserved

Old Ponies vs Popularity & Power

Fig. 9

PowerfulvsPu..	Popula..	PonyPopularityCategory				Grand ..
		Pony	Pariah ..	Null	Number of Records	
Powerful_Pony	30	46		2	1	78
Pony	122	313	27	1	313	463
Puny_Pony	49	130	27	2		208
Grand Total	201	489	54	5		749

- 157 Puny (total – popular) and 54 Pariah ponies do not add substantial value to the worth of farm, and can be removed (esp. those who are both pariah and puny - 27) based on consideration of cost of upkeep vs worth, and effect on worth of others

4. Effects of Food Consumed on Health and Power

Power vs Food

Fig. 10

Treatedo..	FOOD	PowerfulvsPuny		
		Pony	Powe..	Puny..
Healthy (including Ponies with The Power of Gratitude)	Black Mushroom	1,257		
	Broccoli	972	10	224
	Candy Bars	1,012	15	201
	Carrots	946	18	214
	Green Tea	996	9	216
	Growth Hormones	936	15	218
	Jelly	967	11	209
	Pizza	926	13	223
	Pudding	950	7	204
	Strawberries	980	14	227
	Sugar cubes	956	26	185
	Sunshine and Rainb..	884	17	136
Treated	Broccoli	42		24
	Candy Bars	42		29
	Carrots	25		23
	Green Tea	25		31
	Growth Hormones	14		16
	Jelly	42		20
	Pizza	26		25
	Pudding	37		26
	Strawberries	28		23
	Sugar cubes	37		23
	Sunshine and Rainb..	144		104

4.1 Health: Fig. 10, 11a demonstrates that a comparable number of ponies (~1250) are fed a particular food item. However, Fig 11b shows that food type **Sunshine and Rainbows** has highest number of unhealthy ponies who had to be treated. On the other end, **Black Mushroom** (displayed in table) had zero treated cases of ponies. **Growth Hormones** offer a second alternative, though not as successful as Black Mushrooms.

4.2 Power: The attached Fig. 10 indicates that **Black Mushroom** has no associated powerful ponies. In fact, no ponies consuming Black Mushrooms have any powers. Thus, while it is preferable from a health standpoint, it won't increase worth significantly due to lack of powers. **Growth Hormones** however, still show a

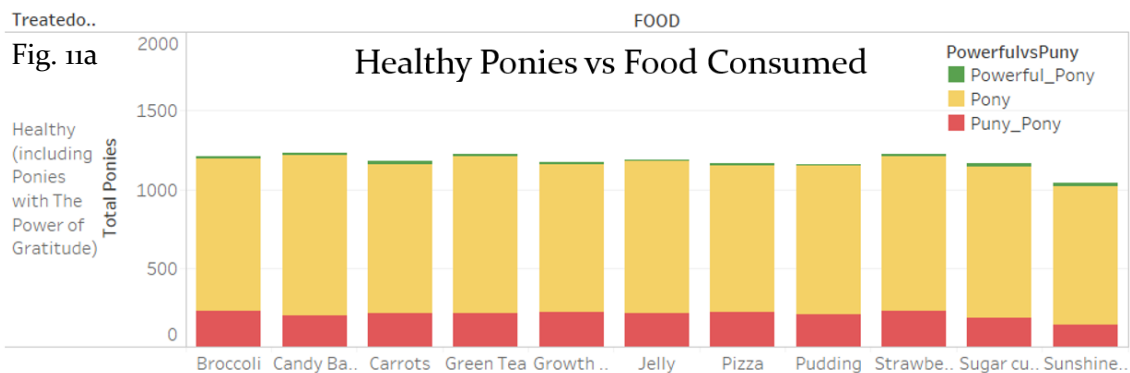
comparable number of powerful ponies (15) to other food types, while providing health benefits. **Sugar Cubes** show a higher number of powerful ponies (26), with comparable health effects.

Thus, I provide three food-based recommendations:

- Avoid using **Sunshine and Rainbows** as food for ponies
- Prefer using **Growth Hormones** or **Sugar Cubes** as food
- Future investigation required on effects of **Black Mushrooms**. It has promising health benefits, but no associated powers are listed.

This should lead to a lower number of treated cases, thus leading to lowered costs or increased market worth of the farm.

Power vs Food

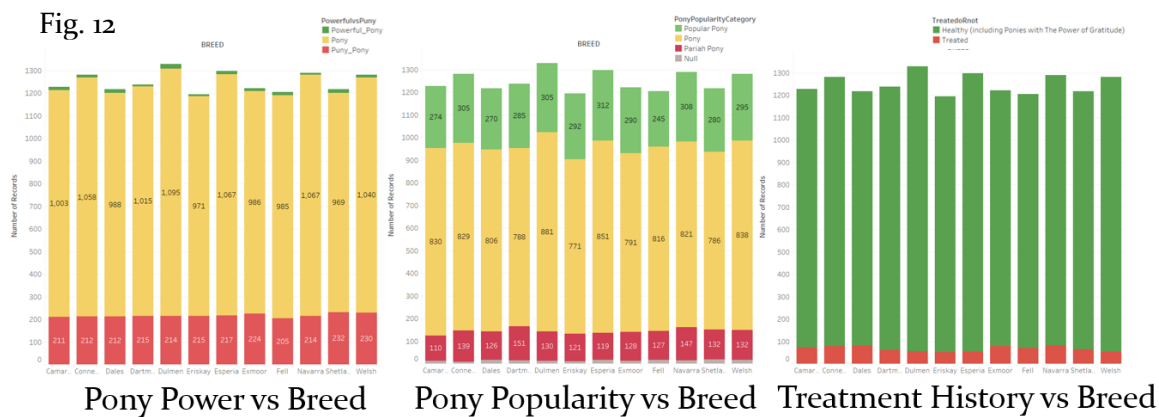


Power vs Food



5. Effect of Breed

Sunshine Valley has 12 breeds of ponies (See Fig. 12). I have looked at the effects of breed on the power of ponies, their popularity, and their treatment history, as illustrated in the charts above. While I do recommend future in-depth analysis of the effects of breed, a preliminary look at the charts (and numbers) indicates **no preference** for a specific breed type for pony health, power or popularity. Thus, it can be ignored for now.



6. Segregation of Ponies into smaller (sub)-farms – based on mutual relationships, popularity, and family members



An extreme step towards improving the worth of Sunshine Valley (which would require future investigation and detailed modeling) is the breakdown of Sunshine Valley into independent sub-farms.

This breakdown will depend on the mutual relationships between ponies, and how it affects their popularity (hence, their worth), along with consideration for family members sticking together when feasible (due to mutual positive effect on worth).

I suggest the following potential sub-farms for investigation:

- Mix popular, powerful ponies (68) with other popular ponies (3,393) on one sub-farm, considering that their popularity remains largely unaffected
 - It could lead to overall increase in value of ponies
- Alternatively, mix ponies based on relationship, with mutual friends in one sub-farm and mutual enemies in separate farms
 - It will positively affect popularity for most ponies, thus creating more value
- All families stick together in nuclear farms with potential family friends around

Provide directions for Future Investigation

Throughout this report, I have identified several cases which require future investigation, which I will repeat here, as well as add other directions:

- a) Investigate the effect of Pariah/Puny Ponies on the popularity of other ponies, and the subsequent impact of removing all of these ponies from the farm
- b) Investigate effect of pony characteristics on power, treatment history & popularity:
 - Proximity of Treated / Sick ponies to each other (identify source)
 - Height / Weight / Breed
- c) Collect data on other pony characteristics for:
 - Selective Breeding of desirable attributes (e.g. power, no past treatments)
 - At what age, does a pony acquire certain Powers? (e.g. if a pony of age >10 is considered powerful, when did it acquire The Power of Gratitude or The Power of Friendship – this would effect popularity, treatment cost)
 - Pony Gender
 - Exercise (and effect on health)
 - Birth vs Death Rate
- d) Investigate the health and power benefits of consuming Black Mushroom (food type) to validate existing data
- e) Investigate the most appropriate approach towards segregation of ponies into smaller farms, leading to maximum worth of farm
- f) To calculate accurate market worth of the farm, I suggest Regression to determine weights for different categories (Power, Popularity, Age etc.), if the final sale value of individual ponies is known
- g) Investigate market factors influencing the volatility of pony market, and demand for older/pariah ponies:
 - Research a substitute use of these undesirable ponies (e.g. pet, track racing, state fairs), apart from sale to toy companies?

Description for SQL Code Components

- Master Code: SQLQueries_for_mastertable_Malvika_Bodh.sql (or .txt)

This file describes the stepwise approach for creating a master table, which is then used for further analysis.

- Final Table: Master_table_Malvika_Bodh.sql (or .txt)

This file includes the final SQL query for creating the table shown below.

PONY_ID	Age	Age_Group	Total_Powers	BREED	FOOD	HEIGHT	WEIGHT	COLOR	PowerfulvsPony	PONY_ID:1	Totalpopularity	PonyPopularityCategory	PONY_ID:2	Treatment	No_ofTreatments
0	15.5057828651421	More than 10	2	Dulmen	Pudding	0.83	184.4	Grey	Pony	0	51	Pony			
1	15.5057828651421	More than 10	1	Connemara	Green Tea	1.09	162.53	Yellow	Pony_Pony	1	87	Pony			
2	15.5057828651421	More than 10	2	Fell	Strawberries	0.78	149.06	Orange	Pony	2	892	Popular Pony			
3	15.503045014355	More than 10	0	Dartmoor	Black Mushroom	0.96	175.83	Orange	Pony	3	136	Pony			
4	15.503045014355	More than 10	1	Dulmen	Growth Hormones	2.56	191.25	Grey	Pony_Pony	4	140	Pony			
5	15.503045014355	More than 10	2	Fell	Candy Bars	0.95	117.59	Yellow	Pony	5	-41	Pariah Pony			

- Analysis: Analysis_Malvika_Bodh.sql (or .txt)

This file is used to group the individual attributes determined in Master_table, and provide actionable insights.