Udacity Project:

Create a Tableau Story

Alan Glasper

16th March 2019

Contents

1 Summary

The story uses the data prepared in the Data Wrangling project with image recognition predictions of what the tweet images contain.

Intended for a non-technical audience, it focuses on the wide range of confidence values of the predictions, using selected We Rate Dogs images as examples.

Of particular interest are examples where the machine reaches high confidence conclusions about the primary object that a human would not.

After demonstrating some reasons for erroneous predictions, it encourages the reader to think about real-life consequences and explore the data themselves to reinforce the points made.

2 Design

In this section I explain the design choices including changes after collecting feedback.

The story prior to feedback (version 1) can be found here:

https://public.tableau.com/views/TheoverconfidentMachinev1/Acautionarytale?:embed=y&:display_count=yes&publish=yes

The story updated from the feedback (final version) can be found here:

https://public.tableau.com/views/OverconfidentMachineStory_v1/Acautionarytale?:embed=y&:display_count=yes&publish=yes

[The links both refer to a v1 because for some reason Tableau Public overwrote the v1 with the final version, even though it had a different name, so I had to upload v1 again].

2.1 Early sketches

First idea was around the concept of identifying what factors lead to higher prediction confidence. Five "chapters" were considered for the structure:

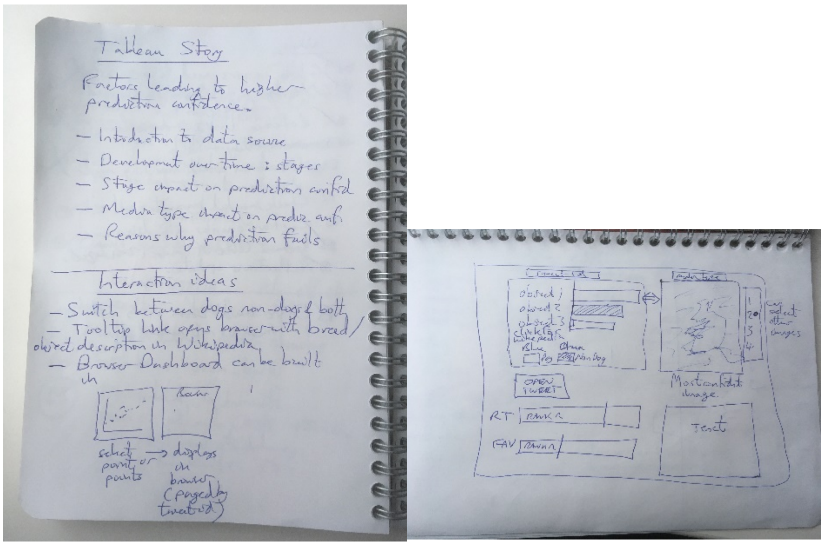
- Introducing the data.
- Development of results over time, also for the dog stages
- The impact on the prediction confidence of the dog stages.
- The impact of the media types.
- Reasons why the prediction fails.

A strong interaction with the data was intended from the beginning:

- Selection of specific dogs to show details and the selected top confidence image.
- Filters to allow different aspects to be explored.
- Animation via paging a variable.
- Dynamic URL actions that would search for information about the predicted object.
- A kind of "dog browser" that would allow the reader to explore the whole dataset.
- Tooltips with summaries of dog information and links for the URL actions.

Some ideas I rejected at this stage were:

- Displaying all the media images for the tweet. This was influenced by a restriction in the version I am using (2018.3) that only allows one web object in a dashboard. (The latest release should allow "zones" to be defined that would allow multiple URLs to be displayed). It was also potentially confusing to the reader so I decided to use only the top confidence image. The other predictions could come from other images than the top one, but the restriction on displaying them meant it was simpler to just not mention it. The focus is anyway on the top prediction.
- Presenting the tweet text. I was considering including the text in the dog browser but decided it distracted from the focus on prediction confidence. Also the data is a little too personal.
- Selecting from scatter plots. I had the idea that selecting a point, e.g. with high retweets and favorites, would open the dog browser. But apart from the really high counts, there is no reason to click in the cloud. I only had the retweet/favorite cloud in mind, so I dropped that.
- Showing the ranking of the retweet and favorite count compared to the others. Again, for the majority in the cloud this is uninteresting and not focused on the main subject.
- Predefined ranges for prediction as categories. This would have introduced more complexity. I dropped all ideas for categories other than the top 5% and bottom 20% of prediction 1.

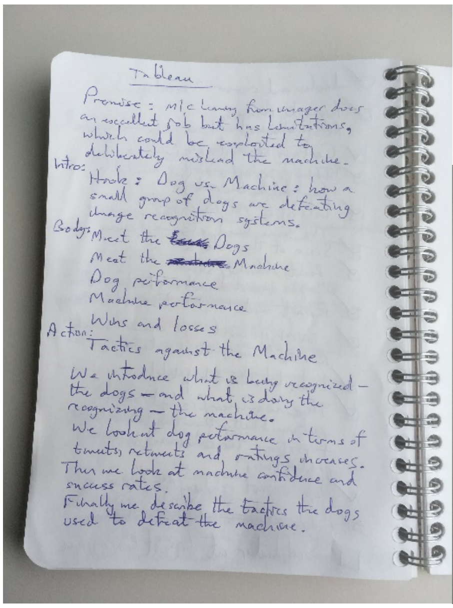


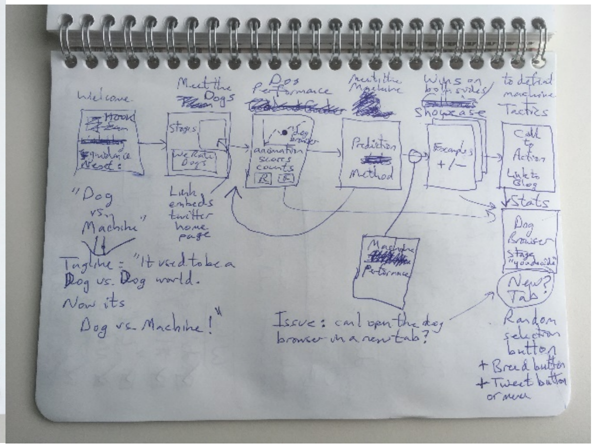
2.2 Refinement of the storyline

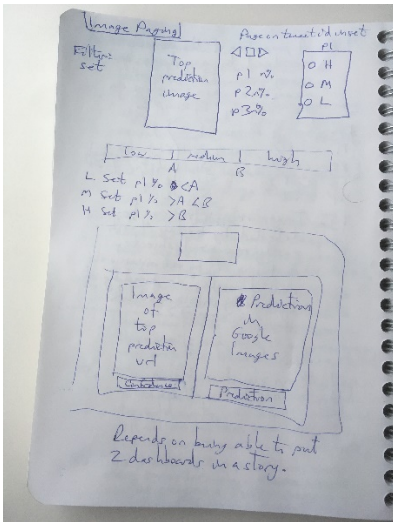
At this stage I was considering framing the story as "Dog vs. Machine", from the dog's viewpoint. As a playful extension of the We Rate Dogs premise, I had ideas to have a dog narrate how, through clever disguises and misdirections, they succeed in fooling the Machine. (This is where the Machine got personified and therefore capitalized, which I stuck with). I would introduce the dogs and the Machine. I would present how the dogs were performing via their ratings and retweets etc., and how the machine was performing via prediction confidence. At the end, their tactics would be presented. I was beginning to create graphics and story elements in Tableau. I had the idea of a Google-style "I'm Feeling Lucky" button that would show a random dog in the browser. The dog image would be shown side by side with the images of the predicted object from Google, mainly because of the exotic dog breeds so that a visual comparison could be made.

Some ideas I rejected at this stage were:

- Having a button to open the tweet. For the same reason as the tweet text, this would have been a distraction and somewhat too personal, although it could have been a solution to how to present all the media images together and would have offered a chance to play the video. This may not have worked from the Tableau Public Server anyway. I did decide to keep the ratings though, since this is the core concept of “We Rate Dogs”.
- Results of analyzing the dog stages, where there were no interesting and reliable results due to the low counts. Again, I decided to keep the stages for one graphic over time, and for the dog information and as a filter in the dog browser.
- Although the framing of dog vs. machine was interesting, I decided it was stretching the situation too far, could be seen negatively, and could give the reader the sense that it is all nonsense. So I opted for a more serious presentation but with fun images and a chance to play.
- The random dog button because it doesn’t really progress the story.
- The side by side presentation of images as I learned that only one URL can be active in one Web component per dashboard, and that it is not possible to show two dashboards in a story element.



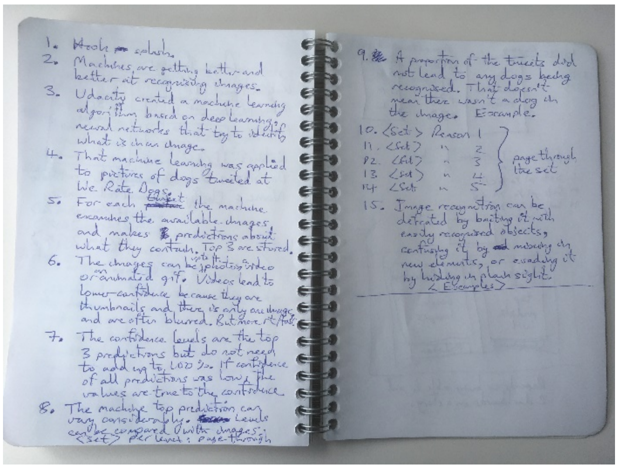




2.3 Structuring the story

At his stage I had enough sense of what would be included to break down the story into plot elements. I wanted to start with a hook that would attract interest. At the end I wanted the reader to take away the idea of how image recognition could be exploited. I had three points in mind: baiting with easily recognized objects like tennis balls, confusing it by mixing in new elements like branches, or hiding in plain sight via disguise.

I had seen some examples of stories where a lot of text is crammed into the Tableau story boxes, which are made large to accommodate it all. I decided on a lightly larger text for readability, and then the box size. Then I wrote the text in short sentences that would move the story along. I would highlight anything interesting in the graphics with short annotations so that I didn't have to explain it in the text box. So the boxes flow the story along in a logical development. I numbered them to facilitate easier feedback, with the intention of removing the numbers at the end. The sequence remained very close to this schema for the rest of the design.

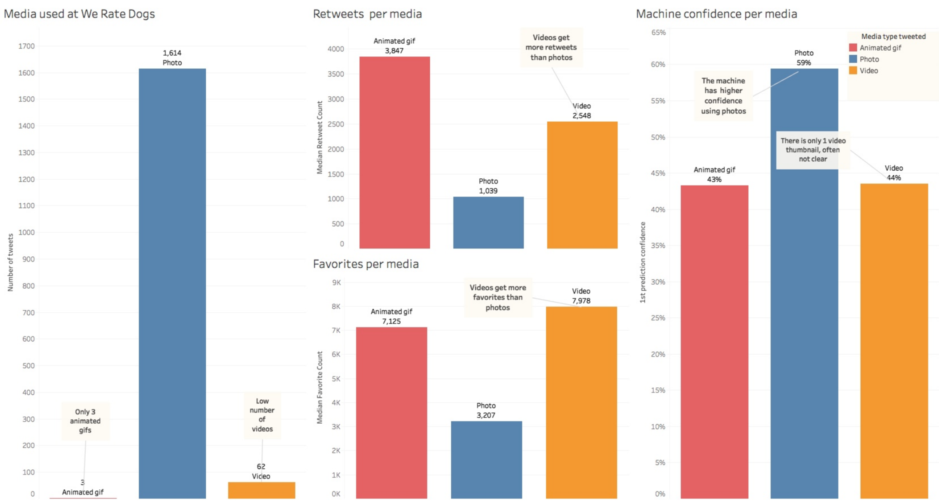


I realized I was not going to be able to open a dog browser window, so I decided to include a text profile of the dog in the tooltip. I came up with the idea that the dog image would update when the reader clicks on a new dog in the tables, and that clicking would also display a menu of links to search for images of predictions 1, 2 or 3. By doing this I could "reuse" the web object in the dashboard. I would have preferred a side-by-side comparison, but this was an acceptable alternative.

I chose a light yellow theme for the story elements and the annotations, so that intuitively background yellow means information. Otherwise it was strictly a low ink-data ratio. Where the x-axis label was obvious, I omitted it. Otherwise labels and titles were included.

Some ideas I rejected at this stage were:

- Paging tweet by tweet. Although it saves space, I decided to present a preselected sample of tweets that would be selected simply by clicking on them. I did consider different ways to present these graphically, but the simplest and best turned out to be a table, with the disadvantage that the tooltip only responds to the final value column. But the readers get to view the prediction names and values easily together. The tooltip would be visible when clicking on a row, and if they hover over or click the final column, which would contain "Full info". Anything else turned out just too complex.
- The following graphic became too complex and I decided that the fact that videos and gifs generate more retweets and favorites to be a no-brainer and not relevant to the story, so they were dropped.
- I experimented with floating legends but it just created more formatting problems, particularly with overlaps with the annotations. So I fixed them at the right of the graphics.
- I removed the dog stage "Multiple" as the number is very small and it would otherwise need explaining. Rather than adding an explanation of the dog stages, I aliased the categories to give a hint about what they mean.

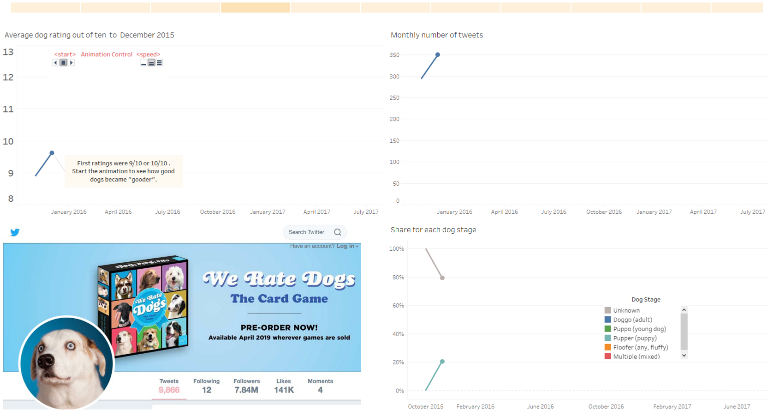


2.4 Final adjustments

By now I had worked through 4 versions of the workbook, adding the screenshot images to round off the story elements.

DogVsMachine_v0.1.twbx	12.03.2019 09:51	Tableau-A...	871 KB
DogVsMachine_v0.2.twbx	13.03.2019 13:39	Tableau-A...	568 KB
DogVsMachine_v0.3.twbx	13.03.2019 15:53	Tableau-A...	746 KB
DogVsMachine_v0.4.twbx	14.03.2019 16:15	Tableau-A...	2.364 KB

I started testing on Tableau Public. I had an animation in v0.4 to get the reader to kick off a graphic, but that had to be dropped because the animations (page shelf) on Tableau Public did not function. It was a nice interaction but not really necessary for the story.



My links to display the object images did not work because Google do not allow embedded use of their pages. A switch to Bing solved that and their image page looks better.

I had to change the size of the image displays in the story element dashboards to be as large as possible, because Tableau Desktop scales the images but Tableau Public does not. So I was getting only parts of the images being displayed, whatever part matched the bitmap size of the web object. The larger size makes the dog images excessively prominent but at least they can be seen without too much scrolling around. I added the tip to scroll if necessary.

3 Feedback

I requested my department manager to provide feedback to version 1. Here is his response. I will respond to the text below.

Antworten Allen antworten Weiterleiten Chat

Fr 15.03.2019 18:39

Michael Horn

AW: Review of Udacity project

An Alan Glasper

Hi Alan,

thanks for this funny story ! It makes me feel more comfortable that humans still have some future before machines are taking over ☺

With regard to your story:

- 1 and 2 are quite clear
- 3 I did not get the story / meaning behind 3 "Udacity". Does it mean Udacity is using ML to create courses or to select courses for individual customers ? What is the context to Dogs, tennis balls and costumes ?
- 4 – who is they ? Maybe you can explain the purpose of applying ML "dog recognition" algorithms (i.e. "idea is to detect dogs, race, age of dog, etc.. or to detect "cute dogs" or other)
- 4 – Average dog rating – Does that mean how users like the dog ? What does 11/10 or 12/10 mean ?
- 4 - shares of Puppy pictures are retweetet more often and then Adults share increases ? What is unknown (the object in the photo or the age).
- 4 - I think the graphs need to be better explained, especially the message behind each graph ...
- 5 – clear, although the # of animated gifs and videos is maybe too low to conclude the confidence level ?
- 6 – this is interactive – maybe there should be a small "manual" for each page to explain what users can do interactively
- 7 – 1st graph – understood – 2nd graph – I need an explanation (but it's late already for me ..)
- 8 – this one I like (including the hint "click one") helps the user to understand the difficulties machines are having in detecting the correct dog
- 9 – does it mean the machine did not find the dog (although it is there) and therefore just identified anything else ? In some cases there is definitely no dog in the picture and the machine identifies another object / animal.
- 9 – the first diagram is not 100% clear to me. Who is ALL ?
- 9 - I'd interpret
  - 1st bar : for 60% of the pictures the machine was confident to have found a dog but only 43% could be named (correctly ?)
  - 2nd bar: 24% of pictures ???
  - 3rd bar: 17% of pictures without a dog –
- 9 – what does dog rating (12/10; 5/10 etc.) mean ?
- 10 – clear
- 11 – strange the machine sees a seat belt, tennis ball or hotdog as primary object when there is clearly a dog on the picture. I did not even see the tennis ball initially (donkey is more present)
- 12 – typo in header (costumes) – other than that clear – if anatomical details of dogs are masked the machine is having difficulties (but not humans – I can still detect dogs)
- 13 – clear - this should improve with more training ?
- 14 – clear, expect the dogsled the prediction is not bad ☺
- 15 – clear
- 16 – maybe include some explanation on how and what can be explored. (not sure every user will now how this works in tableau)

In general: Maybe you can spend some words in the beginning explaining the setup and purpose of this use case / experiment.

Now on to your questions:

1. What do you notice in the visualizations? What stands out?
- Pictures ! (especially for picture recognition use cases as any human can see mistakes made by machines immediately)
- Graphs should be explained / what is the message ?

2. What questions do you have about the data?
- I didn't get which of the data fields are really relevant to the story and message (like dog name – is it relevant ? What does the dog rating mean (12/10) ?

3. What relationships do you notice?
   I think the key message is in 10-15.
   Clarity, Sharpness, In Focus, few or 1 object only, no masking/costumes, High contrast, perspective and object size provide much better results.

4. What do think the main takeaway from this story is?
   Don't trust the machine unless confidence level is very high – but even then be cautious
   The recognition algorithms need more training and refinement, i.e. divide picture in segments in which to search for dogs – then apply identification.
   Maybe you should include some examples for medium confidence level to help understanding the importance of confidence level for proper recognition
   What level of confidence shall be trusted (100%) only
   Could you include a graph showing the percentage of correct recognition vs. confidence level ?

5. Is there something you don't understand from the visualizations?
   See above.

Hope that helps !

Best regards
Michael

3.1 Feedback text

Hi Alan,

thanks for this funny story ! It makes me feel more comfortable that humans still have some future before machines are taking over ☺

With regard to your story:

- 1 and 2 are quite clear
- 3 I did not get the story / meaning behind 3 “Udacity”. Does it mean Udacity is using ML to create courses or to select courses for individual customers ? What is the context to Dogs, tennis balls and costumes ?
- 4 – who is they ? Maybe you can explain the purpose of applying ML “dog recognition” algorithms (i.e. “idea is to detect dogs, race, age of dog, etc.. or to detect “cute dogs” or other)
- 4 – Average dog rating – Does that mean how users like the dog ? What does 11/10 or 12/10 mean ?
- 4 - shares of Puppy pictures are retweetet more often and then Adults share increases ? What is unknown (the object in the photo or the age).
- 4 - I think the graphs need to be better explained, especially the message behind each graph …
- 5 – clear, although the # of animated gifs and videos is maybe too low to conclude the confidence level ?
- 6 – this is interactive – maybe there should be a small “manual” for each page to explain what users can do interactively
- 7 – 1st graph – understood – 2nd graph – I need an explanation (but it’s late already for me ..)
- 8 – this one I like (including the hint “click one”) helps the user to understand the difficulties machines are having in detecting the correct dog
- 9 – does it mean the machine did not find the dog (although it is there) and therefore just identified anything else ? In some cases there is definitely no dog in the picture and the machine identifies another object / animal.
- 9 – the first diagram is not 100% clear to me. Who is ALL ?
- 9 - I’d interpret
    - 1st bar : for 60% of the pictures the machine was confident to have found a dog but only 43% could be named (correctly ?)
    - 2nd bar: 24% of pictures ???
    - 3rd bar: 17% of pictures without a dog –
- 9 – what does dog rating (12/10; 5/10 etc.) mean ?
- 10 – clear
- 11 – strange the machine sees a seat belt, tennis ball or hotdog as primary object when there is clearly a dog on the picture. I did not even see the tennis ball initially (donkey is more present)
- 12 – typo in header (costumes) – other than that clear – if anatomical details of dogs are masked the machine is having difficulties (but not humans – I can still detect dogs)
- 13 – clear - this should improve with more training ?
- 14 – clear, expect the dogsled the prediction is not bad ☺
- 15 – clear
- 16 – maybe include some explanation on how and what can be explored. (not sure every user will now how this works in tableau)

In general: Maybe you can spend some words in the beginning explaining the setup and purpose of this use case / experiment.

Now on to your questions:

1. What do you notice in the visualizations? What stands out?

- Pictures ! (especially for picture recognition use cases as any human can see mistakes made by machines immediately)
- Graphs should be explained / what is the message ?

2. What questions do you have about the data?

- I didn't get which of the data fields are really relevant to the story and message (like dog name – is it relevant ? What does the dog rating mean (12/10) ?

3. What relationships do you notice?
   I think the key message is in 10-15.
   Clarity, Sharpness, In Focus, few or 1 object only, no masking/costumes, High contrast, perspective and object size provide much better results.

4. What do think the main takeaway from this story is?
   Don't trust the machine unless confidence level is very high – but even then be cautious
   The recognition algorithms need more training and refinement, i.e. divide picture in segments in which to search for dogs – then apply identification.
   Maybe you should include some examples for medium confidence level to help understanding the importance of confidence level for proper recognition
   What level of confidence shall be trusted (100%) only
   Could you include a graph showing the percentage of correct recognition vs. confidence level ?

5. Is there something you don't understand from the visualizations?
   See above.

Hope that helps !

Best regards

Michael

3.2 Positive highlights from the feedback

- It is thorough feedback from a trusted source who is known as an excellent communicator.
- The story is entertaining, which was the intention. It can be fun for the reader.
- 8 of the 16 elements were clear as they are.
- Element 8 was especially appreciated, because it "helps the user to understand the difficulties machines are having in detecting the correct dog". Also the hint to "click one".
- Picture examples are a good choice to convey a message about image recognition.
- The key messages were correctly and thoroughly identified.
- The main takeaway was perfectly described, as shown by the suggestion to segment the image. So there is already thinking about how one might resolve the issues.

3.3 Mapping feedback to modifications

In this table I list the feedback items, then decide whether to adopt the feedback ot not. Then I describe the modification if yes, or the reason if no.

Feedback item	Adopt Y/N?	Modification or reason for not adopting feedback
3 I did not get the story / meaning behind 3 "Udacity". Does it mean Udacity is using ML to create courses or to select courses for individual customers ? What is the context to Dogs, tennis balls and costumes ?	Y	Improve the explanation, if necessary add a text box.
4 – who is they ? Maybe you can explain the purpose of applying ML "dog recognition" algorithms (i.e. "idea is to detect dogs, race, age of dog, etc.. or to detect "cute dogs" or other)	Y	Slide 4 needs to be completely reworked. The algorithm is just detecting objects but at a fine level of detail, i.e. not just 'dog' but 'this obscure breed of dog'. I can see that the dog stage is thought to be the goal of the prediction but it was not. Detecting dog cuteness should definitely be Udacity's next project.
4 – Average dog rating – Does that mean how users like the dog ? What does 11/10 or 12/10 mean ?	Y	If I mention the ratings, it has to be seen only in the context of We Rate Dogs, i.e. the core of what they do. It is part of the fun of the story. So I will remove them from the various tables as they are a distraction, and only use them again in the final browser. An extra description will be added to element 4.
4 - shares of Puppy pictures are retweetet more often and then Adults share increases ? What is unknown (the object in the photo or the age).	Y	The easiest solution would be to remove the dog stages altogether as they add little to the story, are low in number and need explanation.
4 - I think the graphs need to be better explained, especially the message behind each graph	Y	Instead of the dog stage graphic, which will be removed, the space will be used for a description that explains the dog rating.
5 – clear, although the # of animated gifs and videos is maybe too low to conclude the confidence level ?	Y	Yes, definitely the animated gifs, as there was only three. But the comparison of video and photo is ok as the difference is large. There is a message about the advantage of more images and the problems with video. But it is a marginal case and is not absolutely essential.
6 – this is interactive – maybe there should be a small "manual" for each page to explain what users can do interactively	Y	Yes, prior to the interactive sections an explanatory element will be added.
7 – 1st graph – understood – 2nd graph – I need an explanation (but it's late already for me ..)	Y	More explanatory text will be added and more interactivity. In addition to the whole graphic options will be added to view the top and bottom 20%. Selecting bars on the histogram will also filter the line chart.
9 – does it mean the machine did not find the dog (although it is there) and therefore just identified	Y	Add more interactivity so that dogs with or without known names can be selected. Link the table to the bar chart so that the connections become clearer.

anything else ? In some cases there is definitely no dog in the picture and the machine identifies another object / animal.		The question made the correct conclusion, hopefully the additional interactivity will remove doubt.
9 – the first diagram is not 100% clear to me. Who is ALL ?	Y	Modify the descriptions to refer to the predictions more clearly.
9 - I'd interpret - 1st bar : for 60% of the pictures the machine was confident to have found a dog but only 43% could be named (correctly ?) - 2nd bar: 24% of pictures ??? - 3rd bar: 17% of pictures without a dog-	Y	This was not understood correctly. Make it clear what having or not having a name means.
9 – what does dog rating (12/10; 5/10 etc.) mean ?	Y	Dog ratings will be removed after element 4 except for the final browser.
12 – typo in header (costumes) – other than that clear – if anatomical details of dogs are masked the machine is having difficulties (but not humans – I can still detect dogs)	Y	Correct typo.
16 – maybe include some explanation on how and what can be explored. (not sure every user will now how this works in tableau)	Y	A user manual will be added earlier, also including the use of filters which is what the comment is referring to. By the time the user reaches here, this will be familiar, as more filters will be added earlier.
In general: Maybe you can spend some words in the beginning explaining the setup and purpose of this use case / experiment.	Y	There will be an introduction in the first elements.
Graphs should be explained / what is the message ?	Y	The problematic graphs will be removed and the remainder made more clear. Still, the intent of the design is that the right conclusions are reached without having to state every point.
I didn't get which of the data fields are really relevant to the story and message (like dog name – is it relevant ? What does the dog rating mean (12/10) ?	Y	The dog names and ratings will be explained earlier.
Maybe you should include some examples for medium confidence level to help understanding the importance of confidence level for proper recognition	N	I did consider having 3 or more bands of confidence, but on examining the images in the middle there was not really a clear distinction from either the top or the bottom of the range. Therefore I chose to focus on the extremes where the message is clearer.
What level of confidence shall be trusted (100%) only	N	This is an excellent question which also shows the message came across. But I don't know a good empirical answer. It depends on the criticality of the decision being made, like all risk assessments. Trying to explain this would add complexity. I am

		happy if the reader goes away asking this very question.
Could you include a graph showing the percentage of correct recognition vs. confidence level ?	N	I did consider at the beginning categorizing all the images for correctness, but I just didn't have the time. This is of course one of the major issues in evaluating such algorithms and again I am happy if the question be asked because the issue was identified. The criteria would have to be laid out in advance and there are several grey areas.

The above modifications were implemented in the final version.

I also requested feedback from the student hub but I didn't receive any feedback suggesting changes.

4 Resources

I referred to the Tableau online documentation frequently. I researched some posts in the Tableau community pages about the restrictions with Tableau Public and potential solutions, related to page shelf animations, resizing of images, and url actions. I was not able to overcome the restrictions but a switch from Google to Bing solved the url action issue.

5 Source Data

The data is the resulting data file from the Data Wrangling project, summarized briefly here:

```
Data columns (total 24 columns):
tweet_id                1680 non-null int64
timestamp               1680 non-null datetime64[ns]
tweet_text              1680 non-null object
rating_numerator        1680 non-null int64
rating_denominator      1680 non-null int64
dog_name                1680 non-null object
rate_type               1680 non-null category
retweet_count           1680 non-null int64
favorite_count          1680 non-null int64
image_urls              1680 non-null object
image_count             1680 non-null int64
image_type              1680 non-null category
expanded_tweet_url      1680 non-null object
url_most_confident_image 1680 non-null object
most_confident_image    1680 non-null category
predict_1_breed         1680 non-null object
predict_1_confidence    1680 non-null float64
predict_1_is_dog        1680 non-null bool
predict_2_breed         1680 non-null object
predict_2_confidence    1680 non-null float64
predict_2_is_dog        1680 non-null bool
predict_3_breed         1680 non-null object
predict_3_confidence    1680 non-null float64
predict_3_is_dog        1680 non-null bool
```

Column	Content
tweet_id	Tweet identifier. Not visible in the story, but used for filters and to define sets of examples.
timestamp	Timestamp. Used for tweet and mean rating history
rating_numerator	Numerator of the dog rating. Used to make a rating string.
rating_denominator	Denominator of the dog rating. Used to make a rating string.
dog_name	Name of dog, also used to create a flag if name is known. Used.
rate_type	Same as Dog Stage. Used in v1, not in final version.
retweet_count	Count of retweets. Used in tooltips.
favorite_count	Count of favorites. Used in tooltips.
image_urls	URLs for the individual images. Not used.
image_count	Count of the individual images. Used in v1, not in final version.
image_type	Referred to as Media Type in the story. Used.
expanded_tweet_url	The full tweet with text: not used.
url_most_confident_image	Image used for the most confident prediction. This is the image that was used in the story.
most_confident_image	The number of the image that gave top confidence. Not used.
predict_1_breed	Predicted object for prediction 1. Same for 2 and 3. Used.
predict_1_confidence	% confidence for prediction 1. Same for 2 and 3. Used.
predict_1_is_dog	Flag whether predicted object 1 was a dog. Same for 2 and 3. Used.