Course 1: How Google does ML

Module 4: Inclusive ML

Lesson Title: **Introduction**

Format: Talking head

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Talent

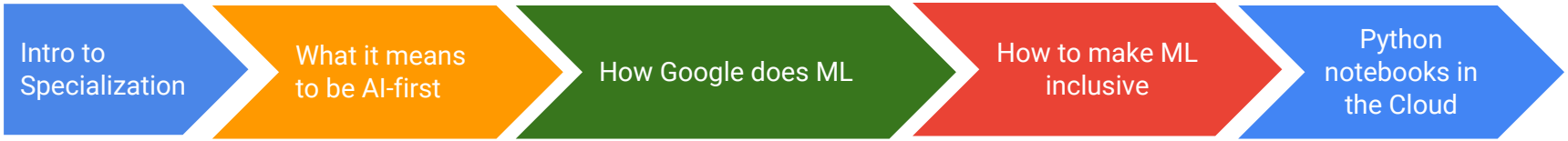
Inclusive Machine Learning

How Google does ML

Machine Learning on Google Cloud Platform

You will learn how to:

- Identify the origins of bias in ML
- Evaluate ML models with biases
- Make models inclusive



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Agenda

Machine learning and human bias

Evaluating metrics with inclusion for your ML system

Equality of opportunity

How to find errors in your dataset using Facets

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Course 1: How Google does ML

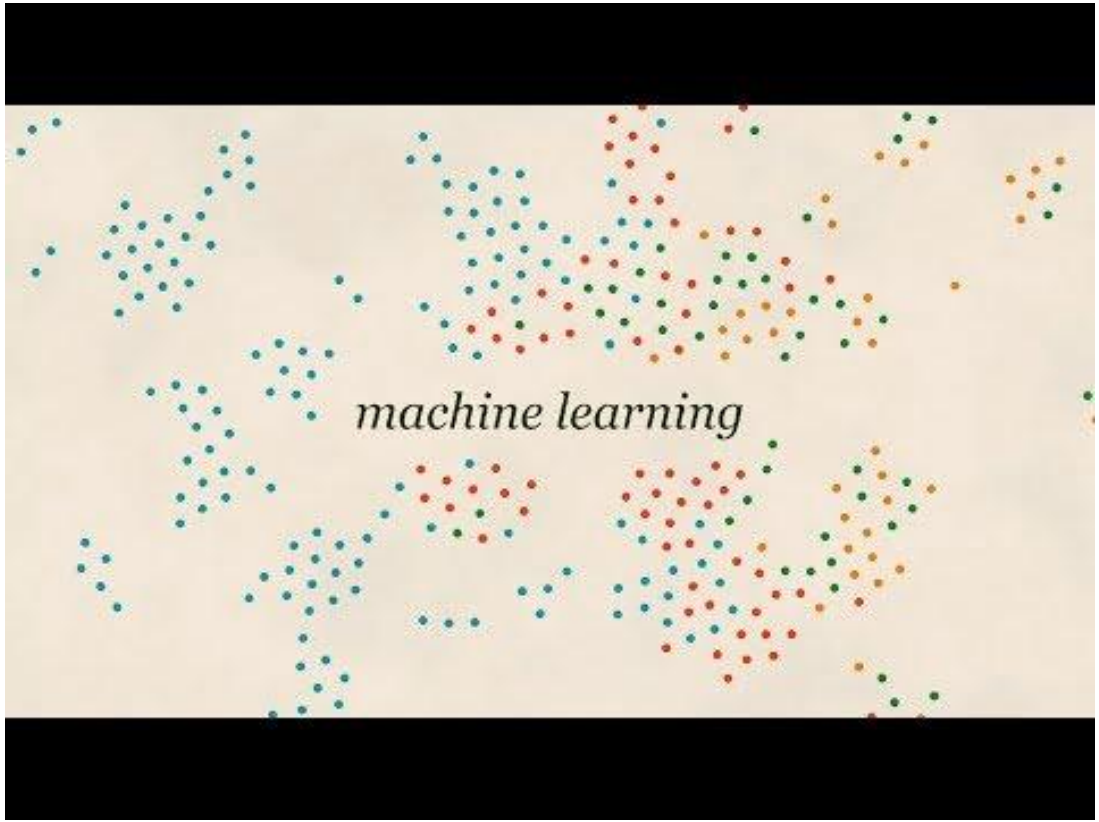
Module 4: How to make ML Inclusive

Lesson Title: **Machine Learning and Human Bias**

Format: Screencast

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Human biases lead to biases in machine learning models



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Course 1: How Google does ML

Module 4: How to Make ML Inclusive

Lesson Title: **Evaluation Metrics Across Subgroups**

Format: Screencast

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

Agenda

Machine learning and human bias

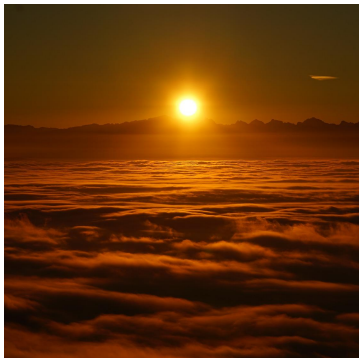
Evaluating metrics with inclusion for your ML system

Equality of opportunity

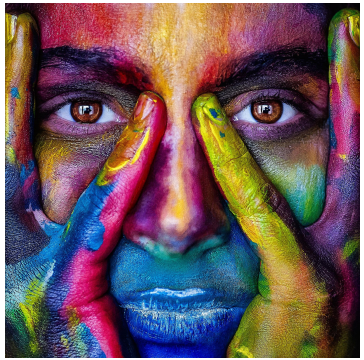
How to find errors in your dataset using Facets

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Evaluate your model over subgroups also








© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

The confusion matrix leads to evaluation metric insights

		Model Predictions: Positive	Model Predictions: Negative
Labels	Positive	**True Positives (TP)** Label says something exists. Model predicts it. 	
Labels	Negative		

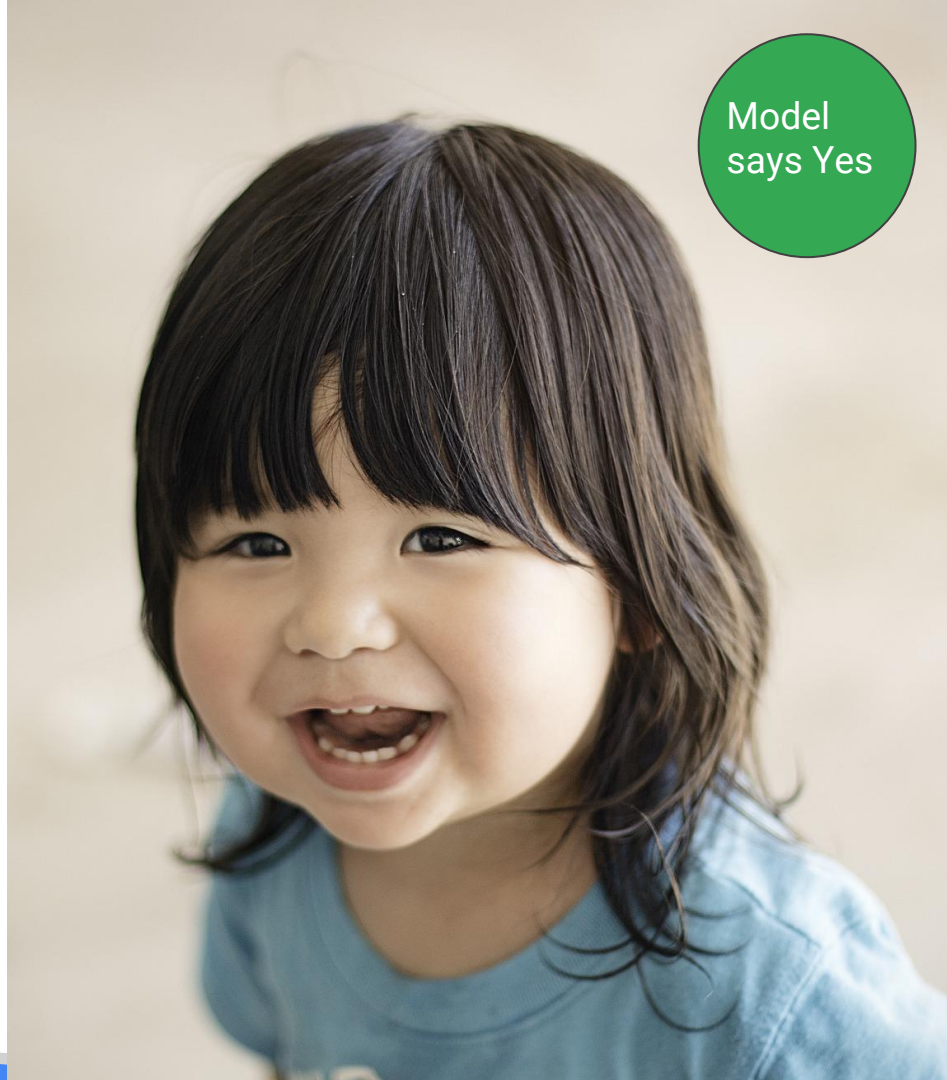
© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

The confusion matrix leads to evaluation metric insights

True Positives (TP)

Label says something exists

Model predicts it exists



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

The confusion matrix leads to evaluation metric insights

		Model Predictions: Positive	Model Predictions: Negative
Labels	Positive	**True Positives (TP)** Something exists. Model predicts it. 	**False Negatives (FN)** *Type II Error* Something exists Model *doesn't* predict it 
	Negative		

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

The confusion matrix leads to evaluation metric insights

False Negatives (FN)

Type II Error

Label says something exists

Model predicts it doesn’t exist

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

The confusion matrix leads to evaluation metric insights

		Model Predictions: Positive	Model Predictions: Negative
Labels	Positive	**True Positives (TP)** Label = something exists. Model predicts it. 	**False Negatives (FN)** *Type II Error* Label = something exists Model *doesn't* predict it 
	Negative		**True Negatives (TN)** Something *doesn't* exist Model *doesn't* predict it 

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

The confusion matrix leads to evaluation metric insights

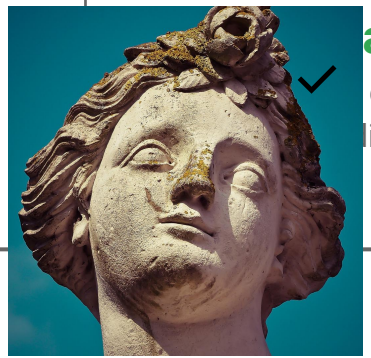
True Negative (TP)

Label says something doesn’t exist

Model predicts it doesn’t exist

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

The confusion matrix leads to evaluation metric insights

		Model Predictions: Positive	Model Predictions: Negative
Labels	Positive	**True Positives (TP)** Something exists. Model predicts it. 	**False Negatives (FN)** *Type II Error* Something exists. Model *doesn't* predict it 
	Negative	**False Positives (FP)** *Type I Error* Something *doesn't* exist Model predicts it 	**atives (TN)** exist lict it 

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

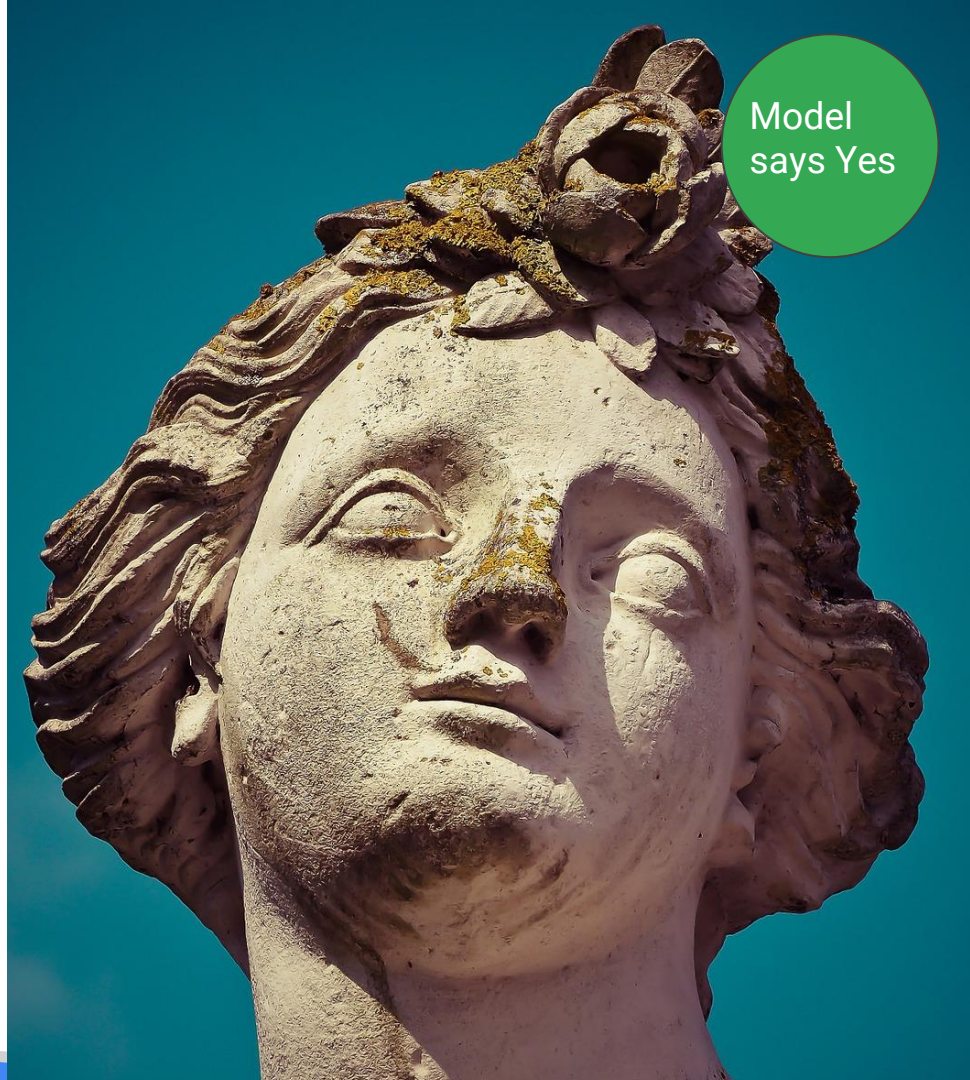
The confusion matrix leads to evaluation metric insights

False Positives (FN)

Type I Error

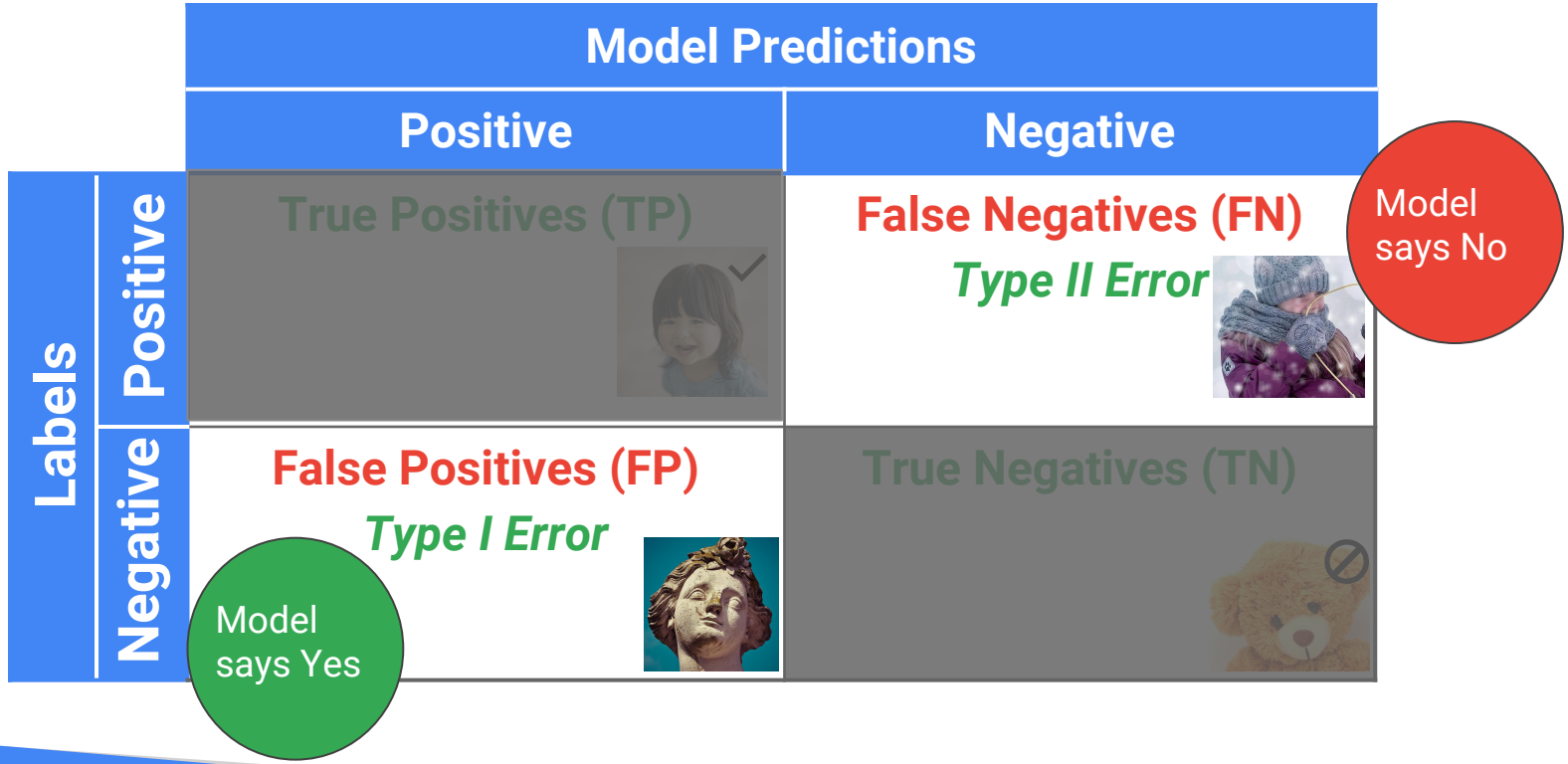
Label says something doesn't exist

Model predicts it exists



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

False positives and false negatives errors occur when predictions and labels disagree



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

Course 1: How Google does ML

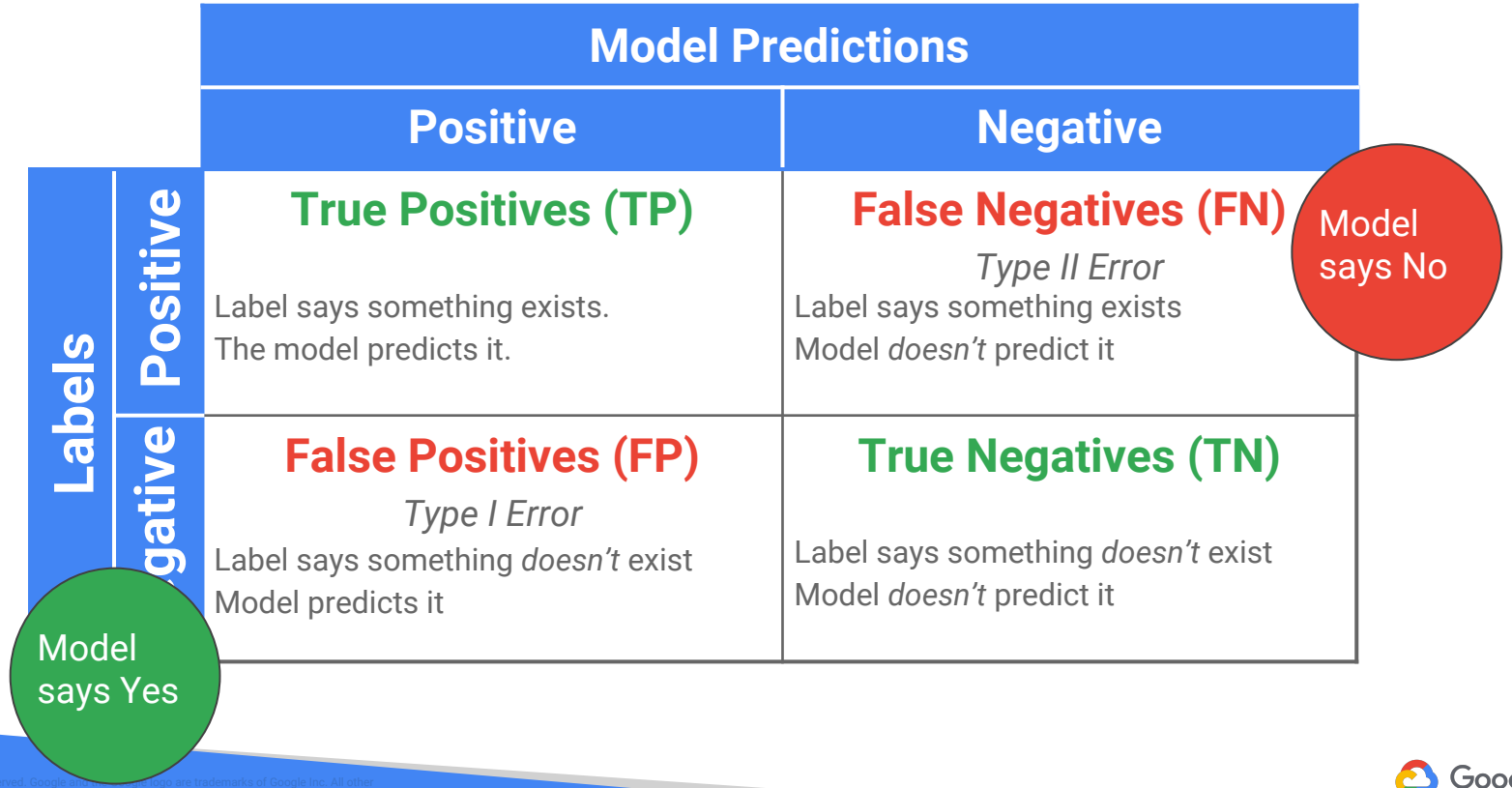
Module 4: Inclusive ML

Lesson Title: **Statistical Measurements**

Format: Screencast

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Evaluation metrics can help highlight areas where machine learning could be more inclusive



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

False negative rate is the fraction of true faces that are not detected by the ML system

		Model Predictions: Positive	Model Predictions: Negative
Labels	Positive	**True Positives (TP)** Label says something exists. The model predicts it.	**False Negatives (FN)** *Type II Error* Label says something exists Model *doesn't* predict it

$$\text{False Negative Rate} = \frac{\text{False Negatives}}{\text{False Negatives + True Positives}}$$

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

False negative rate is the fraction of true faces that are not detected by the ML system

True Positives (TP)

Label says something exists.
The model predicts it.

False Negatives (FN)

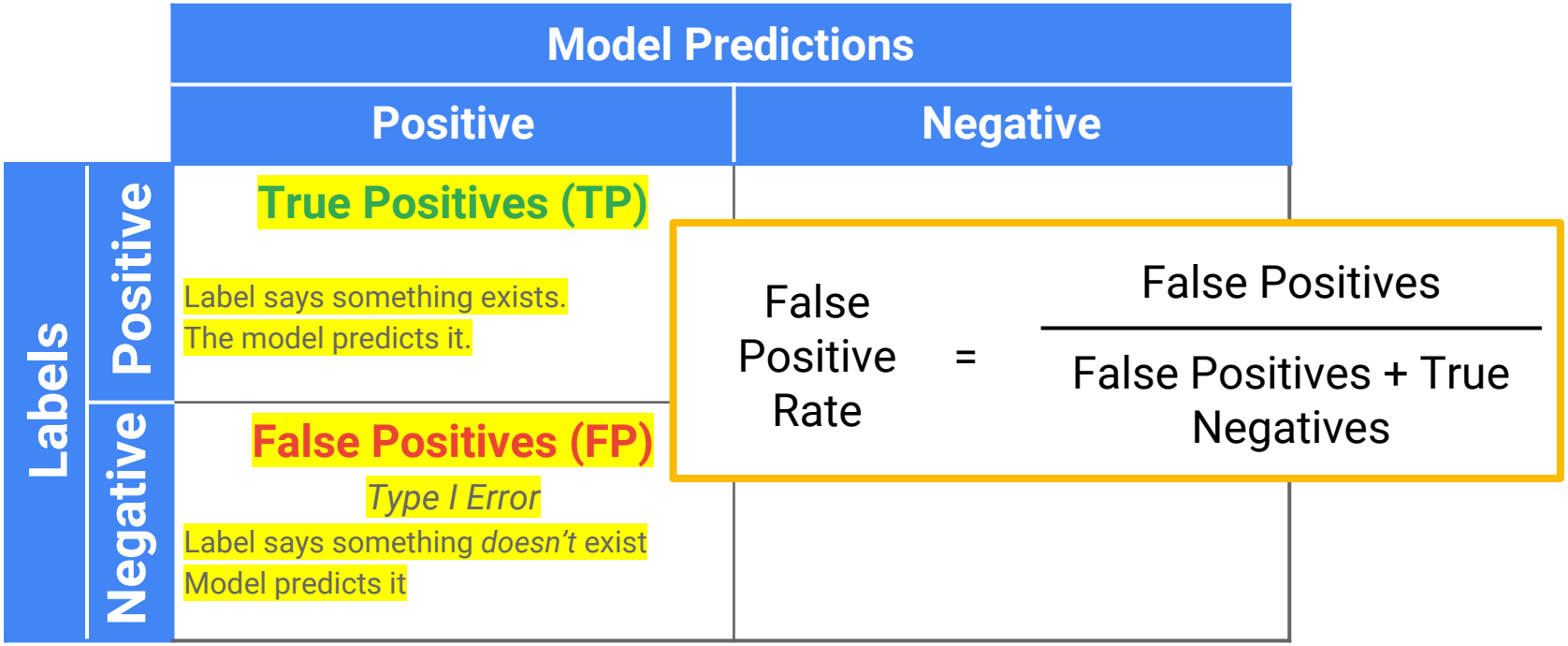
Type II Error

Label says something exists
Model *doesn't* predict it

$$\text{False Negative Rate} = \frac{\text{False Negatives}}{\text{False Negatives + True Positives}}$$

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

False positive rate is the fraction of the faces that the ML model detects that are not really faces



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

False positive rate is the fraction of the faces that the ML model detects that are not really faces

True Positives (TP)

Label says something exists.
The model predicts it.

False Positives (FP)

Type I Error

Label says something *doesn't* exist
Model predicts it

$$\text{False Positive Rate} = \frac{\text{False Positives}}{\text{False Positives + True Negatives}}$$

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Sometimes false positives are better than false negatives

Privacy in images

False positive

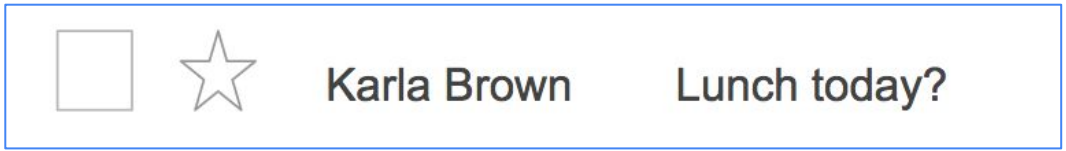
False negative

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Sometimes false negatives are better than false positives

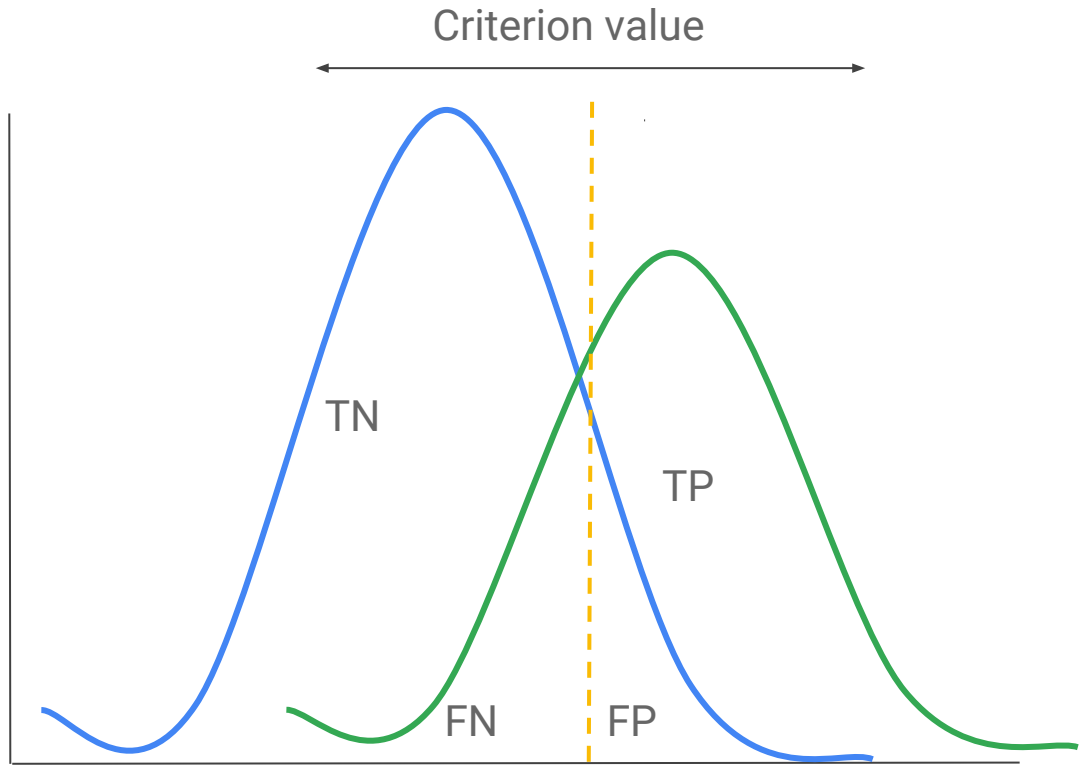
False Negative: E-mail that is SPAM is not caught, so you see it in your inbox.

False Positive: E-mail flagged as SPAM is removed from your inbox.



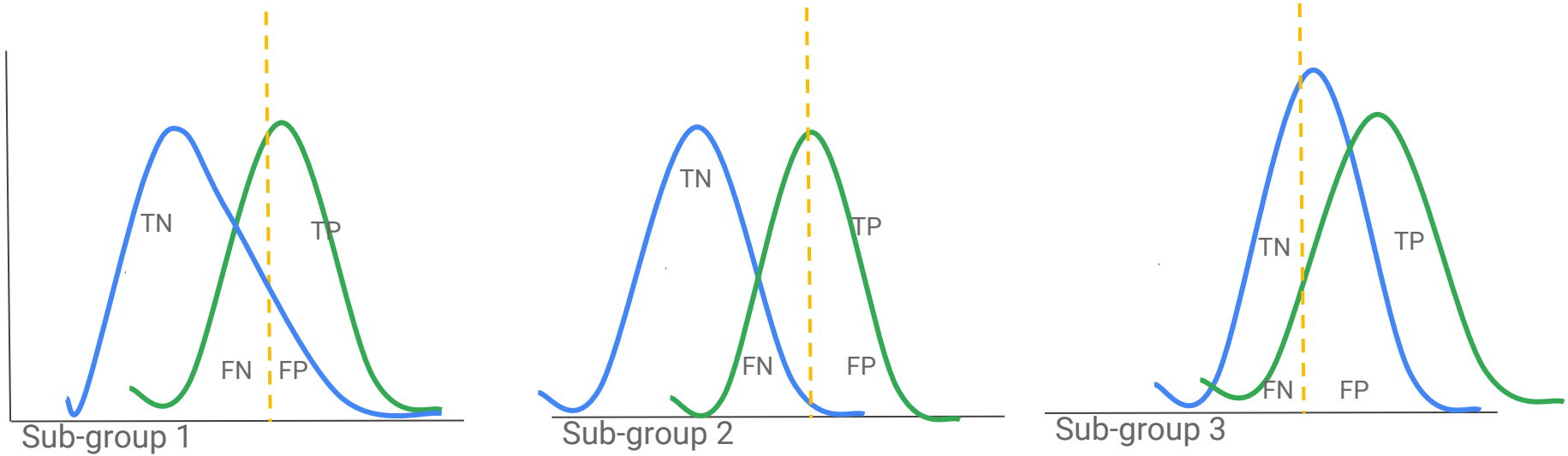
© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Find the threshold that brings the precision or recall to acceptable values



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Check the precision/recall you obtain with that threshold in each of your subgroups



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.


Choose your evaluation metrics in light
of acceptable tradeoffs between
False Positives and False Negatives.

Course 1: How Google does ML

Module 4: Human-centered ML

Lesson Title: **Equality of Opportunity**

Format: Screencast

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

Agenda

Machine learning and human bias

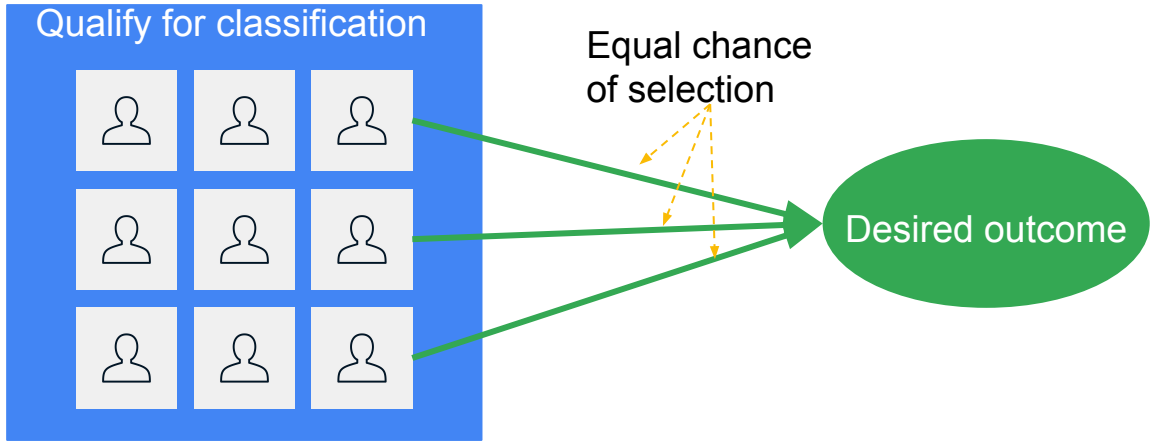
Evaluating metrics with inclusion for your ML system

Equality of opportunity

How to find errors in your dataset using Facets

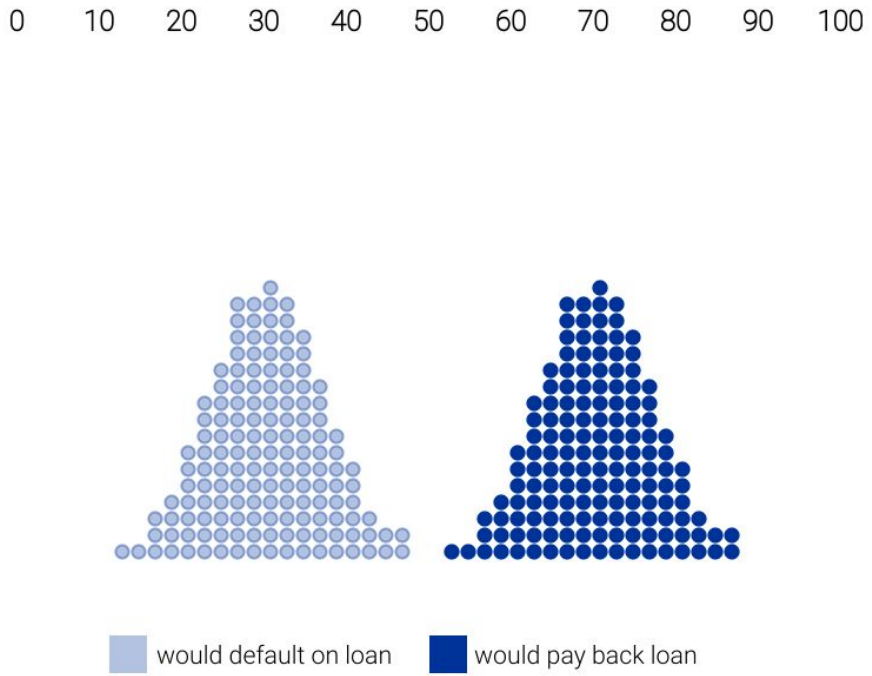
© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

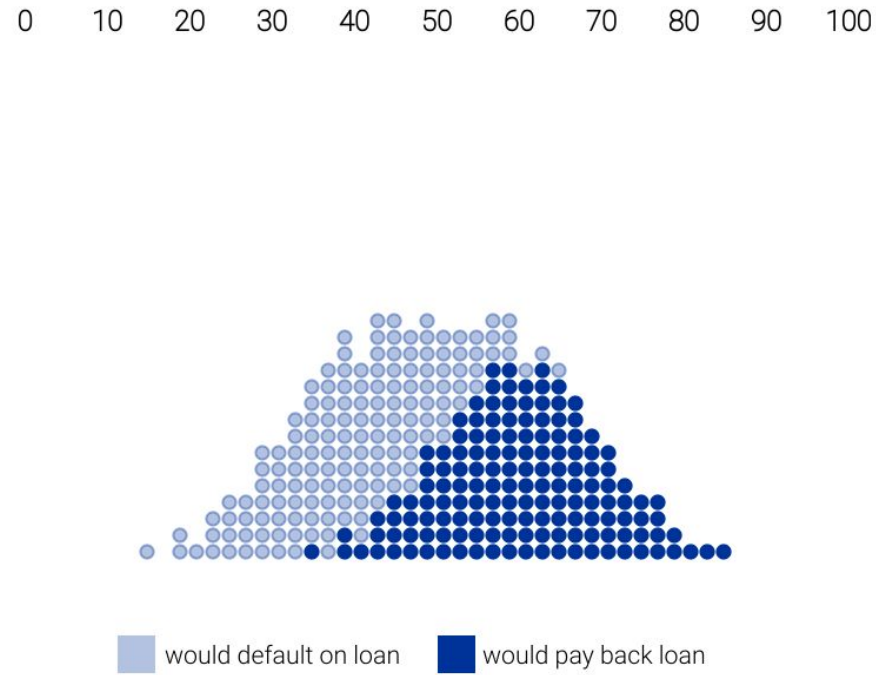
The Equality of Opportunity approach strives to give individuals an equal chance of desired outcome



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

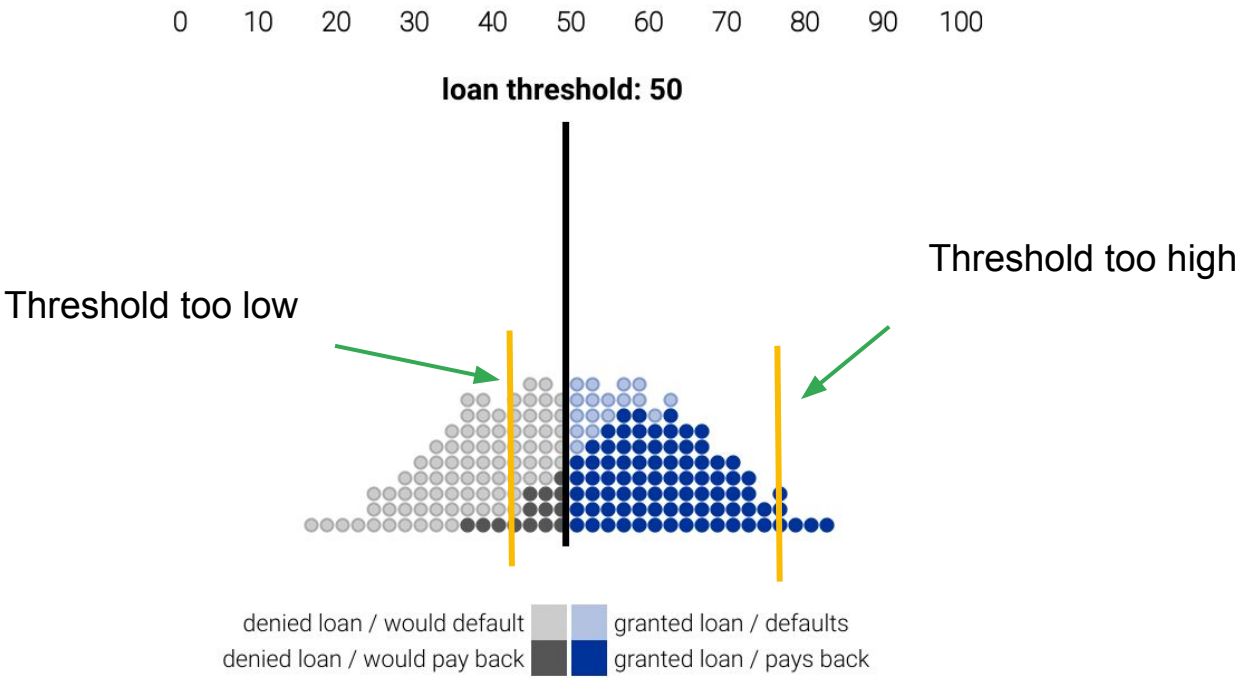
A toy classifier to predict who will pay back their loan involves two populations that might overlap





© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

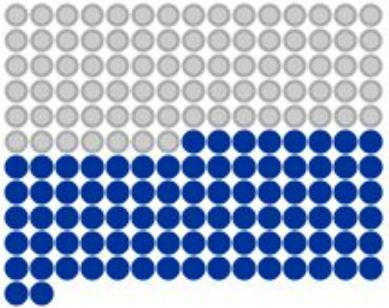
Picking a credit score threshold involves a tradeoff



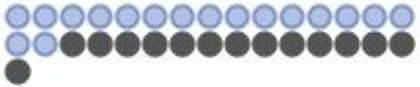
© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

The impact of a threshold on credit score is evaluated based on its impact on customers and on loan repayment

Correct 84%
loans granted to paying applicants and denied to defaulters



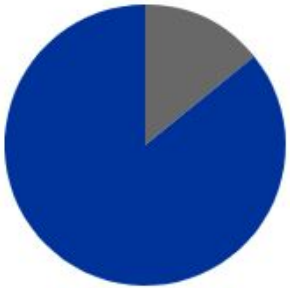
Incorrect 16%
loans denied to paying applicants and granted to defaulters



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

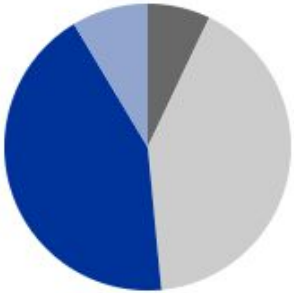
Simulating the impact of a threshold on profit

True Positive Rate 86%
percentage of paying applications getting loans



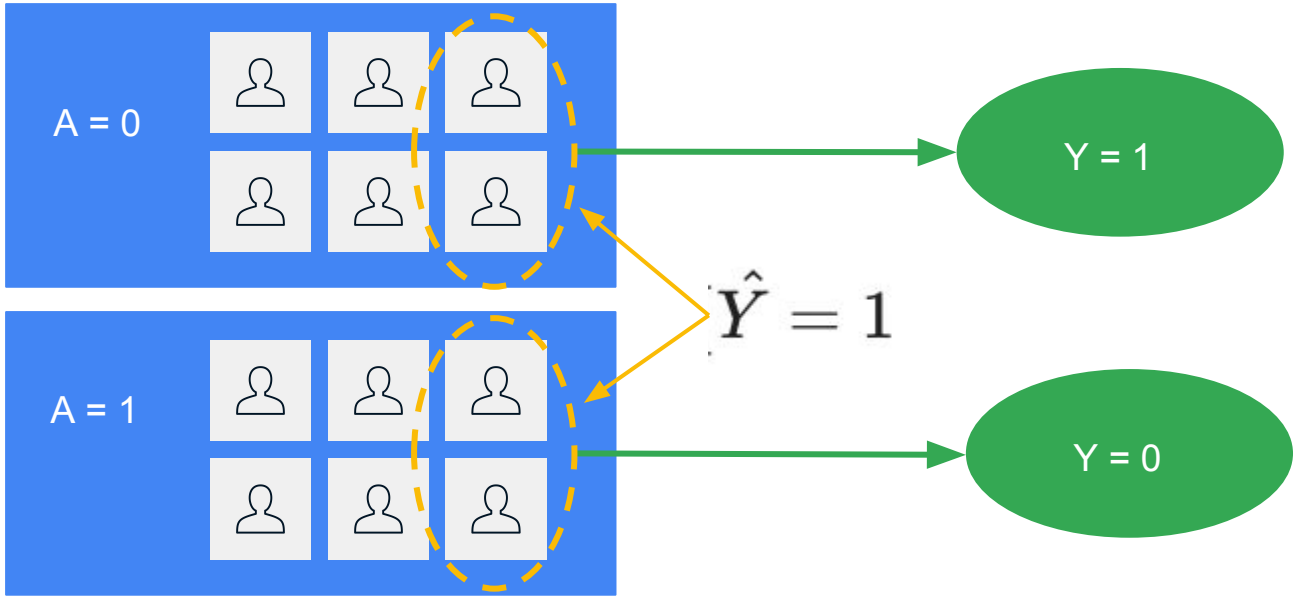
Profit: **13600**

Positive Rate 52%
percentage of all applications getting loans



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Classification and Discrimination must obey the Equality of Opportunity Principle



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Course 1: How Google does ML

Module 4: Human-centered ML

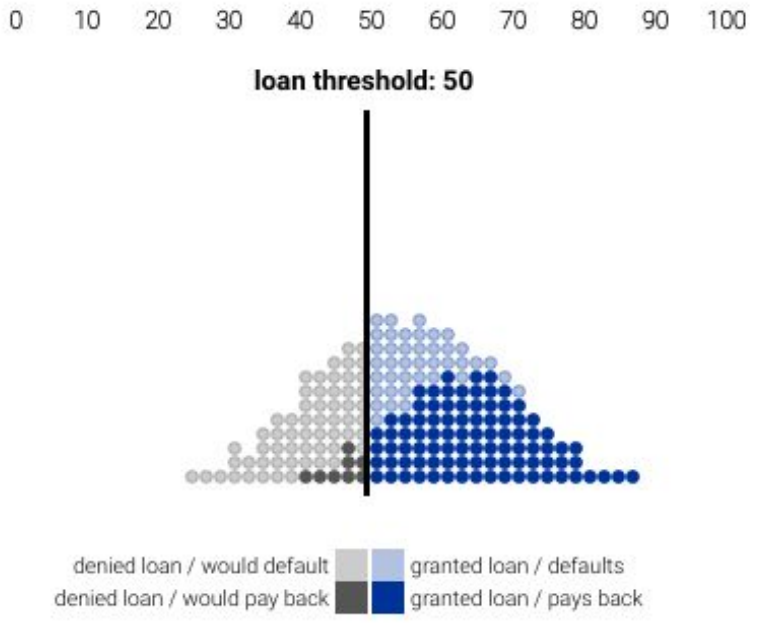
Lesson Title: **Simulating Decisions**

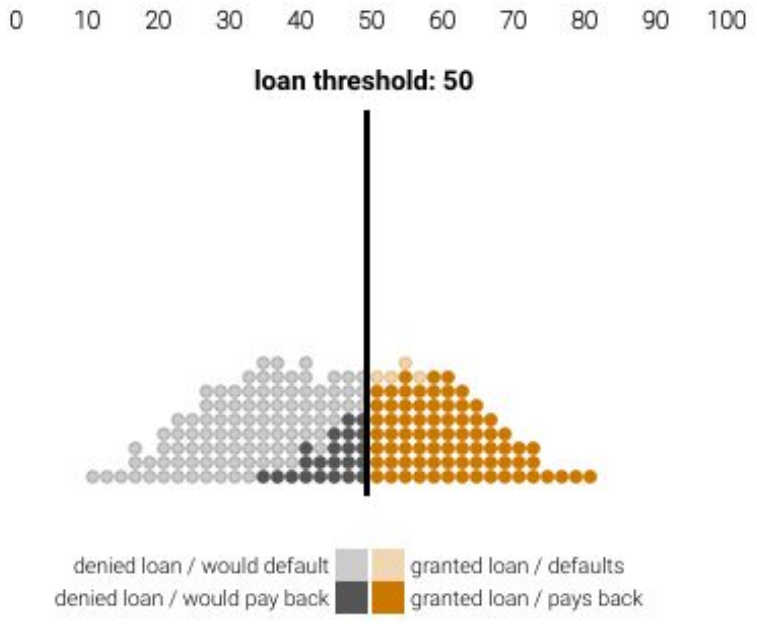
Format: Screencast

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

Simulating decisions with no constraints can lead to unequal distribution





A successful loan makes $300
An unsuccessful loan costs $700
Credit scores are between 0 - 100

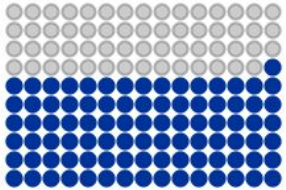
© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

Simulating decisions with no constraints can lead to unequal distribution

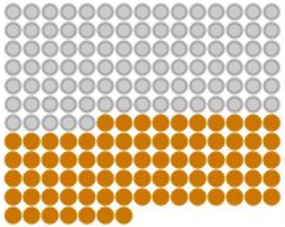
Total profit = **19600**

Correct 76%
loans granted to paying applicants and denied to defaulters



Incorrect 24%
loans denied to paying applicants and granted to defaulters

Correct 87%
loans granted to paying applicants and denied to defaulters



Incorrect 13%
loans denied to paying applicants and granted to defaulters

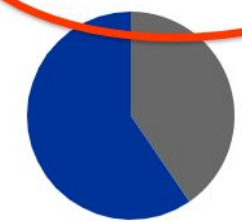
Threshold

- Credit Score of 50 for Blue Group
- Credit Score of 50 for Orange Group

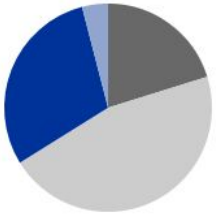
© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Simulating decisions for max profit result in unequal standards

True Positive Rate 60%
percentage of paying applications getting loans

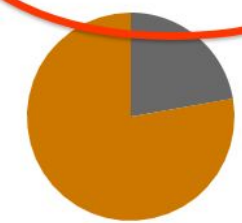


Positive Rate 34%
percentage of all applications getting loans



Profit: **12100**

True Positive Rate 78%
percentage of paying applications getting loans



Positive Rate 41%
percentage of all applications getting loans

Profit: **20300**

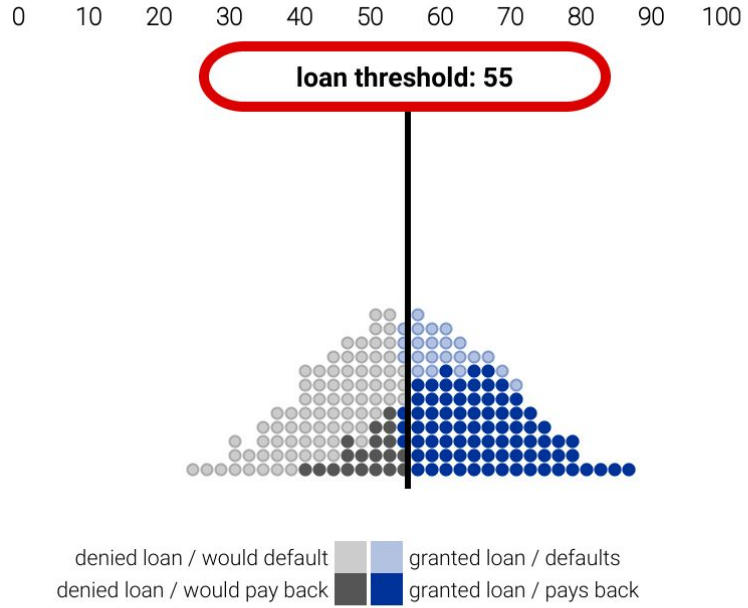
Threshold

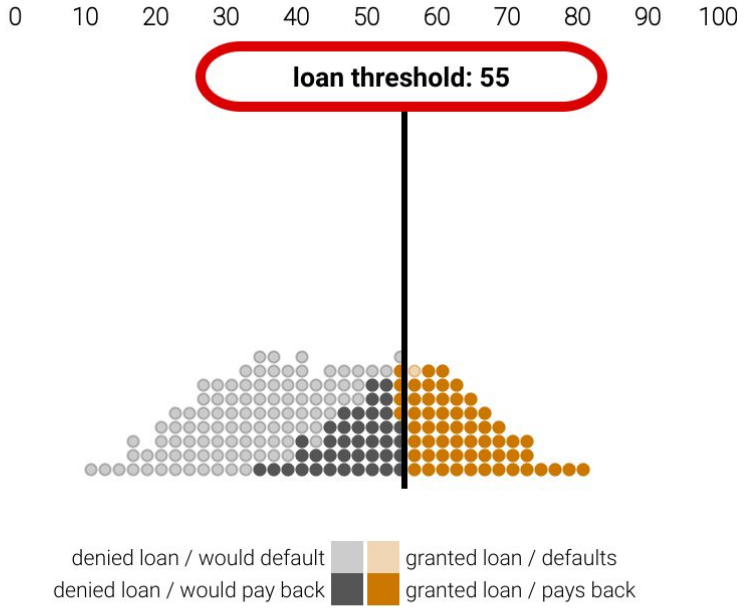
- Credit Score of 61 for Blue Group
- Credit Score of 50 for Orange Group

Total profit: **32400**

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Simulating decisions with group unaware holds everyone to the same standard, which can be unfair to some groups



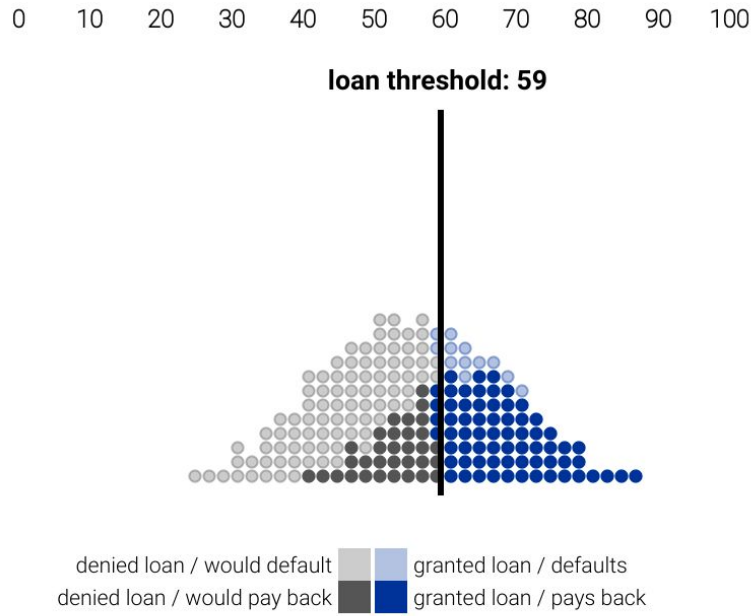


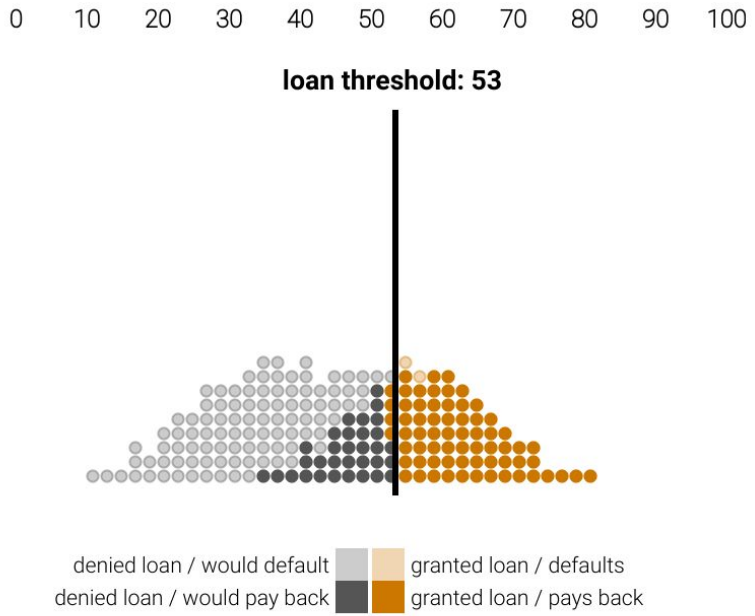
Total profit = **25600**

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

Simulating decisions equal opportunity results in an identical true positive rate for all groups





Total profit = **30400**

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

Course 1: How Google does ML

Module 4: Human-centered ML

Lesson Title: **Finding Errors in Your Dataset Using Facets**

Format: Screencast

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

Agenda

Machine learning and human bias

Evaluating metrics with inclusion for your ML system

Equality of opportunity

How to find errors in your dataset using Facets

https://research.googleblog.com/2017/07/facets-open-source-visualization-tool.html

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Facets gives users a quick understanding of the distribution of values across features of their datasets

Sort by

Non-uniformity ☐ Reverse order Feature search

Features: ☑ int(6) ☐ string(9)

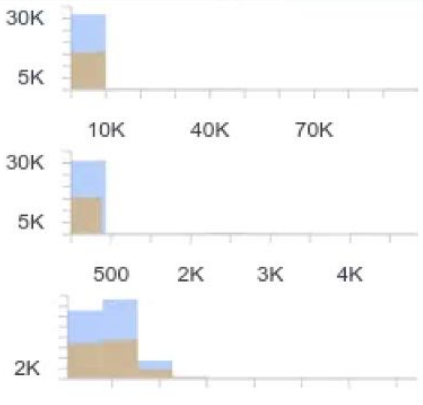
train test

Numeric Features (6)

Chart to show: Standard

☐ log ☐ expand ☐ percentages

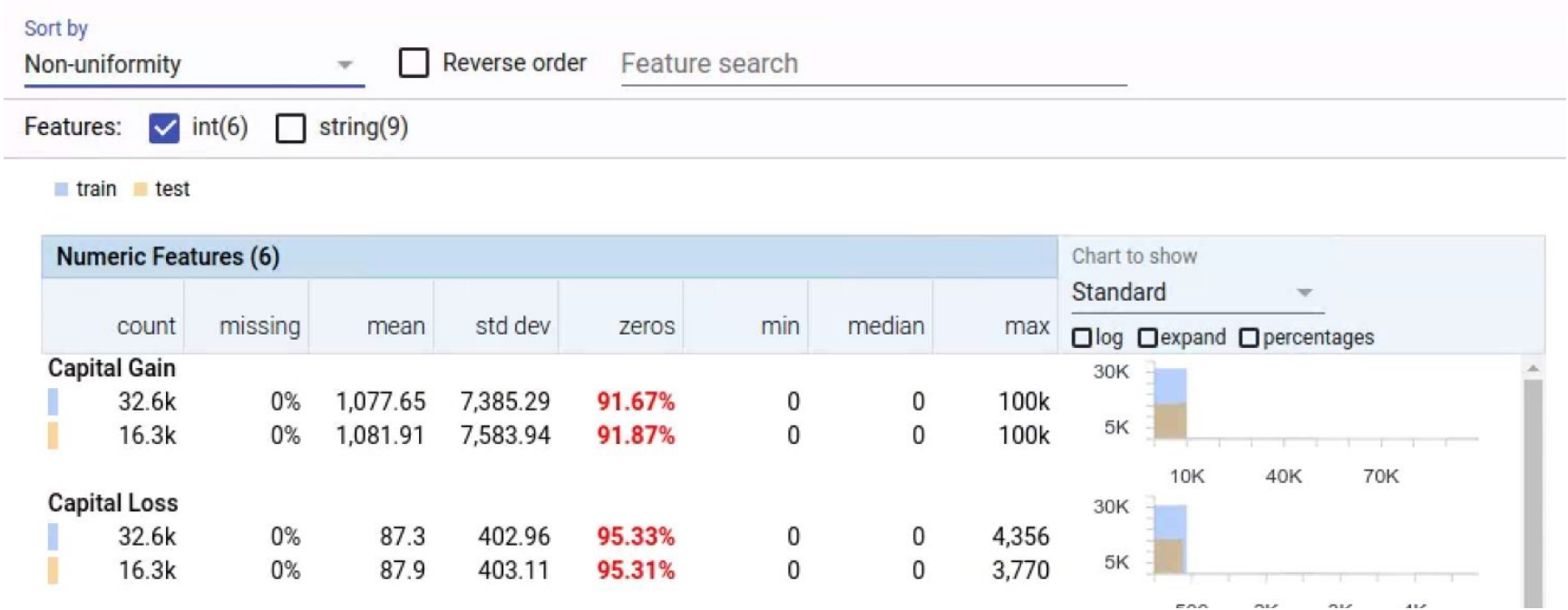
	count	missing	mean	std dev	zeros	min	median	max
Capital Gain								
train	32.6k	0%	1,077.65	7,385.29	91.67%	0	0	100k
test	16.3k	0%	1,081.91	7,583.94	91.87%	0	0	100k
Capital Loss								
train	32.6k	0%	87.3	402.96	95.33%	0	0	4,356
test	16.3k	0%	87.9	403.11	95.31%	0	0	3,770
fnlwgt								
train	32.6k	0%	190k	106k	0%	12.3k	178k	1.48M
test	16.3k	0%	189k	106k	0%	13.5k	178k	1.49M



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

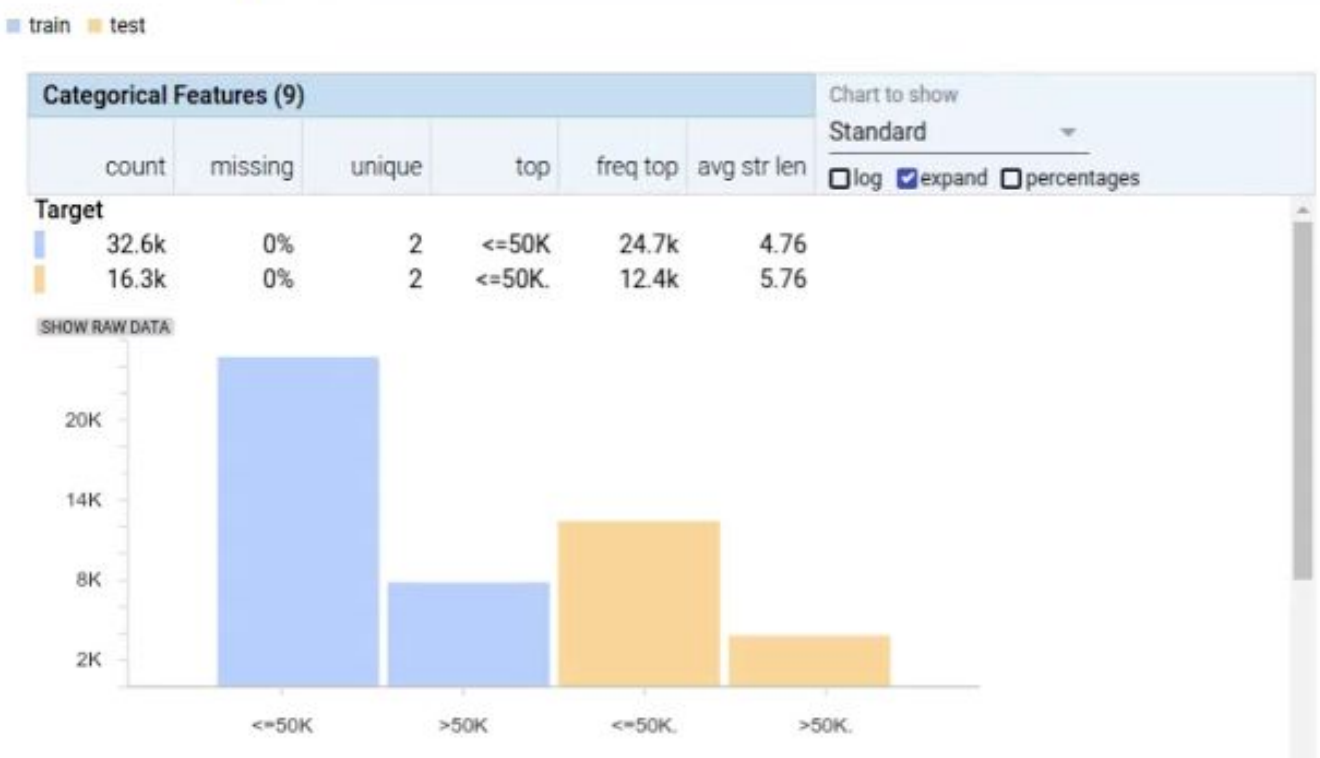
In Facets features are sorted by non-uniformity, with the feature with the most non-uniform distribution at the top



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

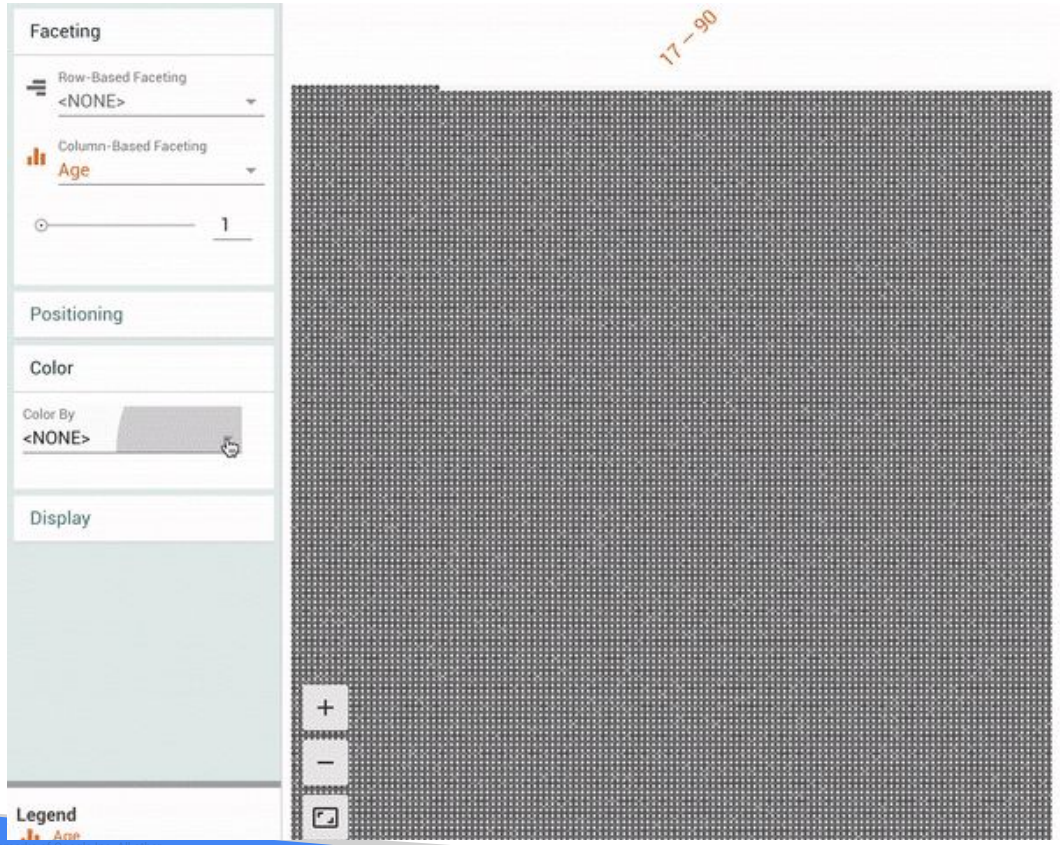
Facets features are sorted by distribution distance



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

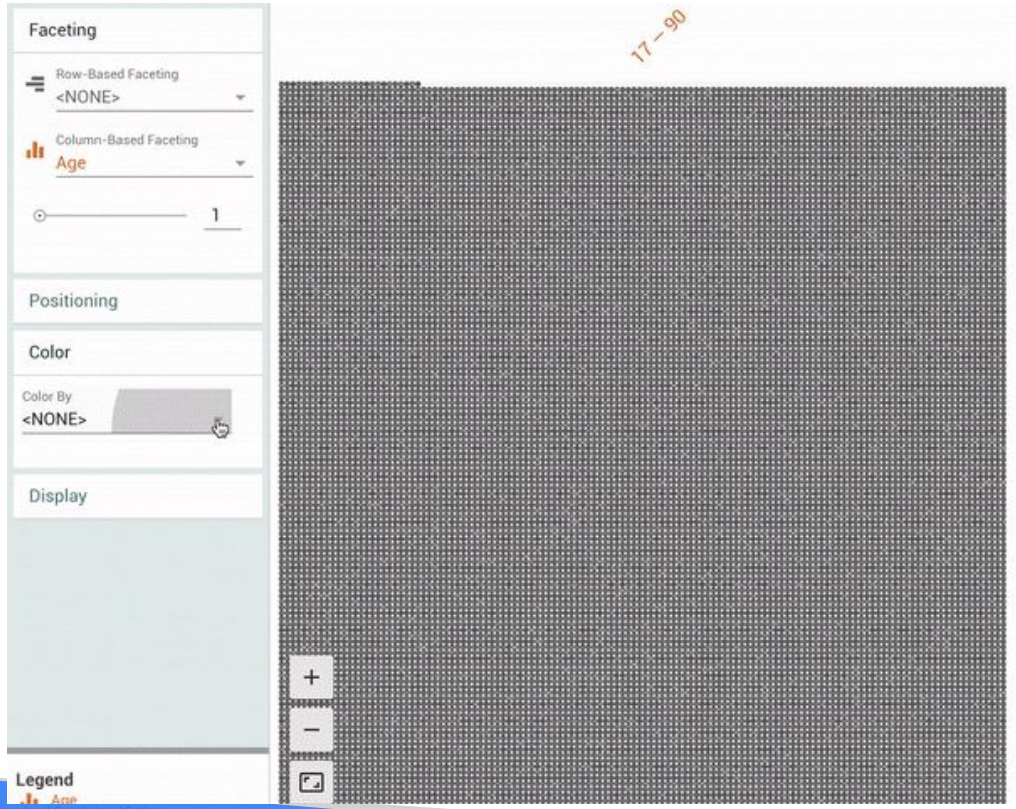
Facets Dive provides an easy-to-customize, intuitive interface



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

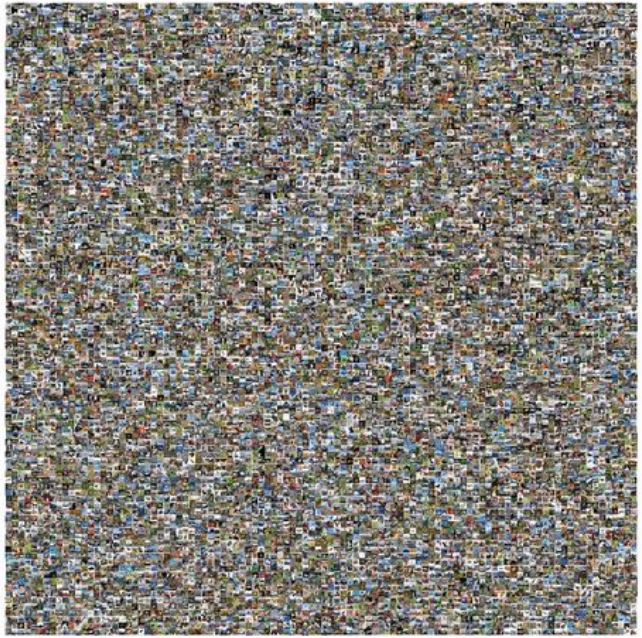
Color the data points by one feature, then facet in another



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud

Explore CIFAR-10 for Errors using Facets Dive



© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Exploring a dataset for Errors

© 2017 Google Inc. All rights reserved. Google and the Google logo are trademarks of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.

Google Cloud


cloud.google.com
Images by Connie Zhou