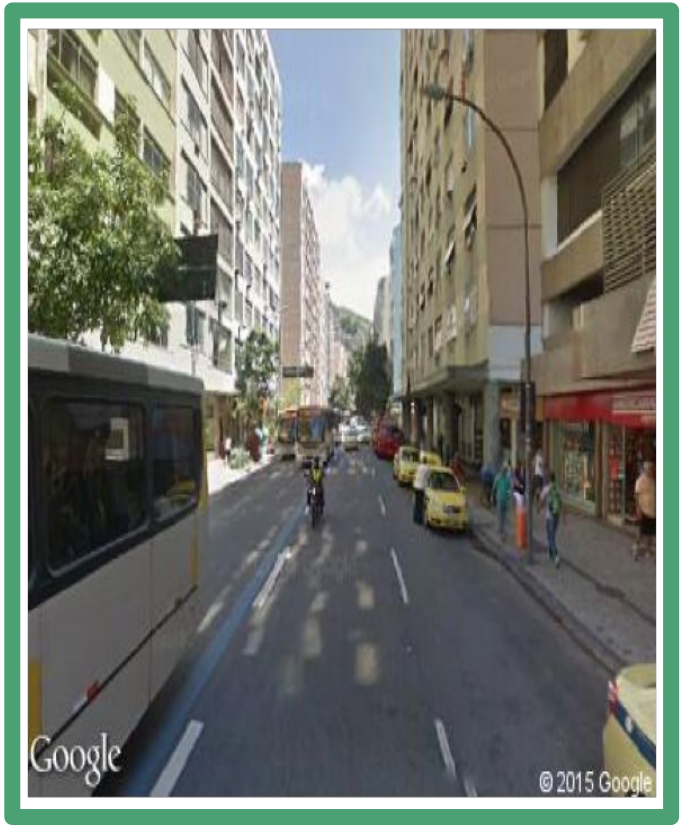
Quantifying Urban Safety Perception on Street View Images

Felipe A. Moreno-Vera, Bahram Lavi, Jorge Poco

Motivation

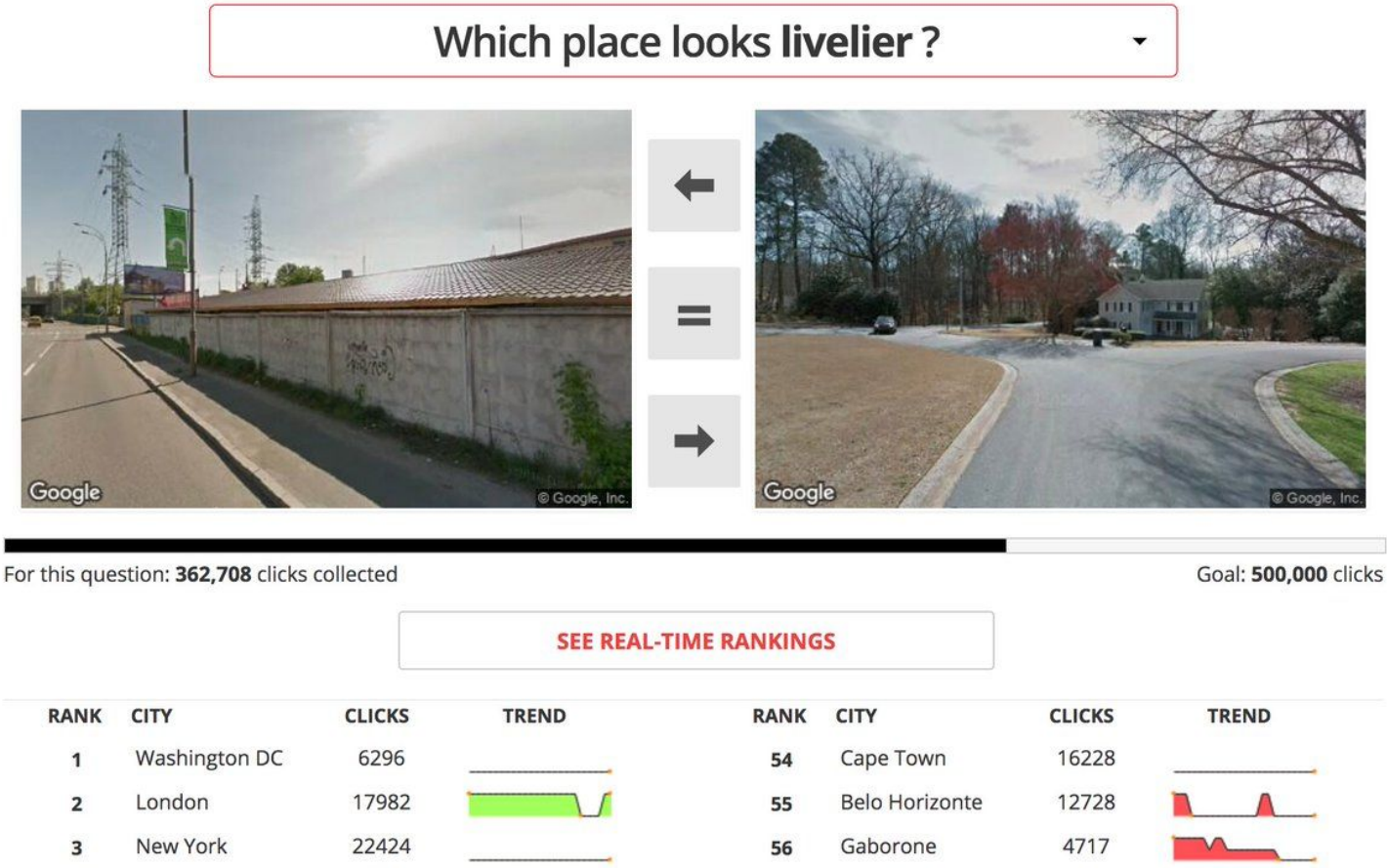
Which one looks safer?

Bangú (RJ)



City Center (RJ)

Place Pulse



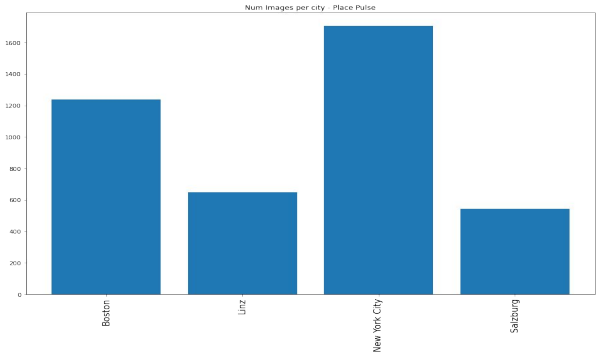
http://pulse.media.mit.edu/

* Comparisons were made using two random images from random cities.

Place Pulse Dataset

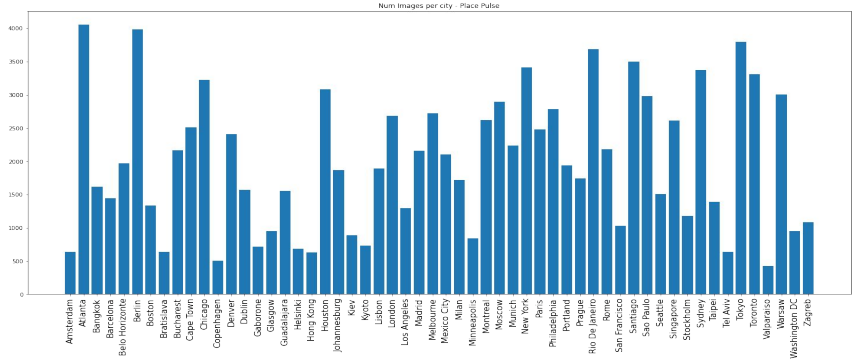
Place Pulse 1.0:

- 73 806 Comparisons, 4 136 images
- 2 Countries (US y Austria)
- 4 cities: New York City, Boston, Linz and Salzburg
- 3 categories: Safe, Wealth and Unique



Place Pulse 2.0:

- 1 223 649 Comparisons, 111 390 images
- 32 countries
- 56 cities
- 6 categories: Safe, Wealth, Depress, Beautiful, Boring, and Lively



* **Remember:** We will focus in **Place Pulse 2.0** only.

Data Pre-processing

Dataset sample: Set of comparisons*

left_id	right_id	winner	left_lat	left_long	right_lat	right_long	category
513d7e23fdc9f	513d7ac3fdc9f	equal	40.744156	-73.93557	-33.52638	-70.591309	depressing
513f320cfdc9f	513cc3acfdc9f	left	52.551685	13.416548	29.76381	-95.394621	safety
513e5dc3fdc9f	5140d960fdc9f	right	48.878382	2.403116	53.32932	-6.231007	lively

* **Remember:** Comparisons were made using two random images from random cities.

Pre-processing Comparisons

Perceptual Scores Approach

Salesse et. al, “The Collaborative Image of The City: Mapping the Inequality of Urban Perception”, 2013

$$W_i = \frac{w_i}{w_i + d_i + l_i}$$

$$L_i = \frac{l_i}{w_i + d_i + l_i}$$

$$q_{i,k} = \frac{10}{3}(W_{i,k} + \frac{1}{n^w_{i,k}}(\sum_{j_1} W_{j_1,k}) - \frac{1}{n^l_{i,k}}(\sum_{j_2} L_{j_2,k}) + 1)$$

*Nassar et al, “The evaluative image of the city”, 1990

Rank Images Approach

Dubey et. al, “Deep Learning the City : Quantifying Urban Perception At A Global Scale”, 2016

$$\mu_x \longleftarrow \mu_x + \frac{\sigma_x^2}{c} \cdot f\left(\frac{(\mu_x - \mu_y)}{c}, \frac{\varepsilon}{c}\right)$$

$$\mu_y \longleftarrow \mu_y - \frac{\sigma_y^2}{c} \cdot f\left(\frac{(\mu_x - \mu_y)}{c}, \frac{\varepsilon}{c}\right)$$

$$\sigma_x^2 \longleftarrow \sigma_x^2 \cdot \left[1 - \frac{\sigma_x^2}{c} \cdot g\left(\frac{(\mu_x - \mu_y)}{c}, \frac{\varepsilon}{c}\right)\right]$$

$$\sigma_y^2 \longleftarrow \sigma_y^2 \cdot \left[1 - \frac{\sigma_y^2}{c} \cdot g\left(\frac{(\mu_x - \mu_y)}{c}, \frac{\varepsilon}{c}\right)\right]$$

$$c^2 = 2\beta^2 + \sigma_x^2 + \sigma_y^2$$

$$q_{i,k} = \frac{10}{c_{max,k}}(c_{i,k})$$

**Minka et al, “TrueSkill 2: An improved Bayesian skill rating system”, 2018

Perceptual Score Approach

$$W_i = \frac{w_i}{w_i + d_i + l_i} \qquad L_i = \frac{l_i}{w_i + d_i + l_i}$$

$$q_{i,k} = \frac{10}{3}(W_i + \frac{1}{w_i}(\sum_{k_1=1}^{w_i} V_w(k_1)) - \frac{1}{l_i}(\sum_{k_2=1}^{l_i} V_l(k_2)) + 1)$$

Salesse et. al, “The Collaborative Image of The City: Mapping the Inequality of Urban Perception”, 2013

*Nassar et al, “The evaluative image of the city”, 1990

Processed sample: Images from Rio de Janeiro - Place Pulse 2.0

Image	ID	Safety	Lively	Wealthy	Beauty	Boring	Depressive
	513d7e23fdc9f	7.42	8.58	6.5	7.3	2.64	1.23
	513f320cfdc9f	6.07	4.97	7.13	8.61	1.67	0.86

* **Note:** We perform the calculation in all categories, but we will focus in safety only.

Dataset Statistics: summary

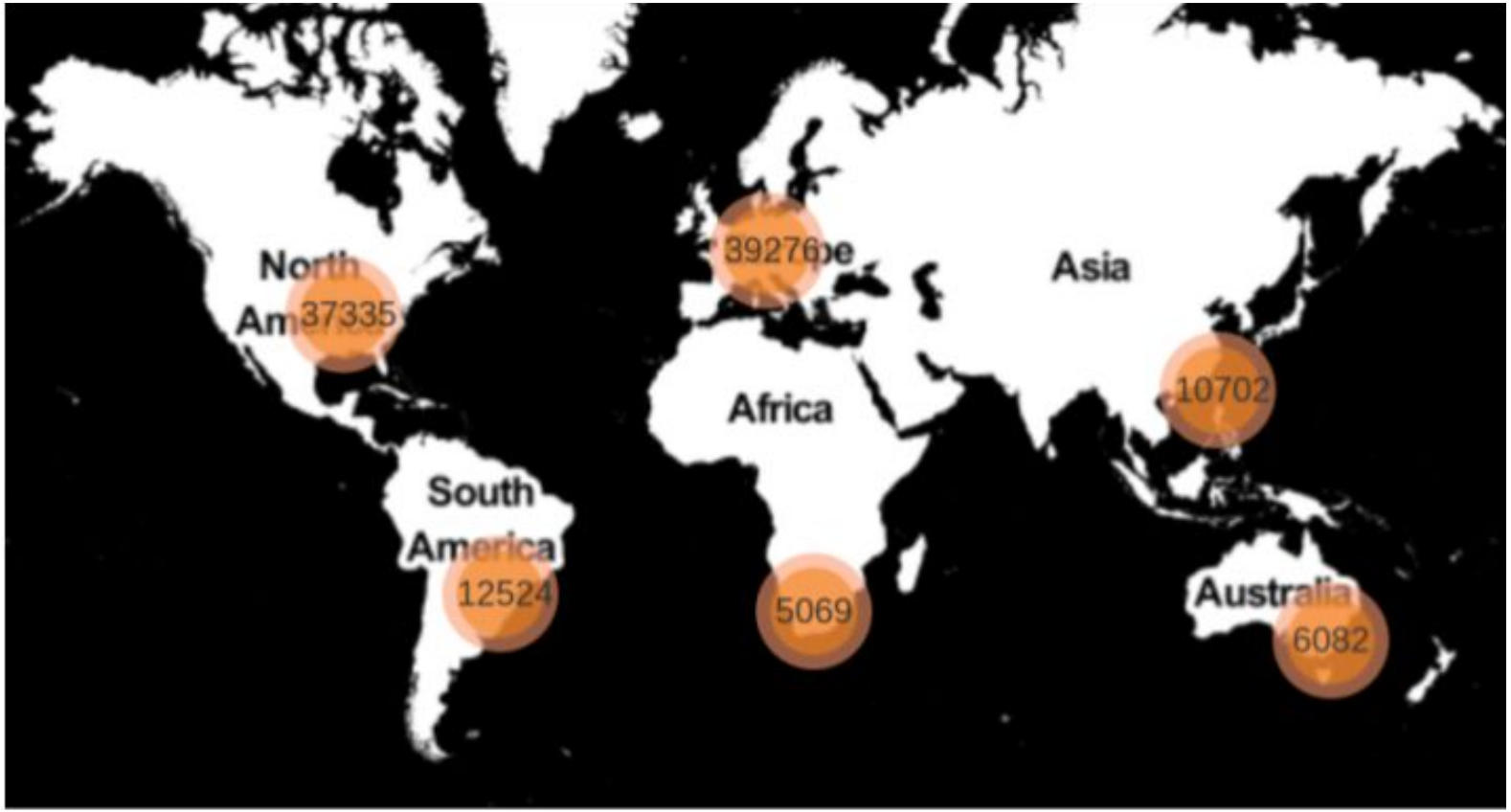
Place Pulse 1.0				
City	# images	*safe mean*	*wealth mean*	*unique mean*
Linz	650	4.85	5.01	4.83
Boston	1237	4.93	4.97	4.76
New York	1705	4.47	4.31	4.46
Salzburg	544	4.75	4.89	5.04
Total	4136			

Place Pulse 2.0			
Continent	#countries	#cities	#images
Europe	19	22	38,747
North America	3	17	37504
South America	2	5	12,524
Asia	5	7	11,417
Oceania	1	2	6,097
Africa	2	3	5,101
Total	32	56	111,390

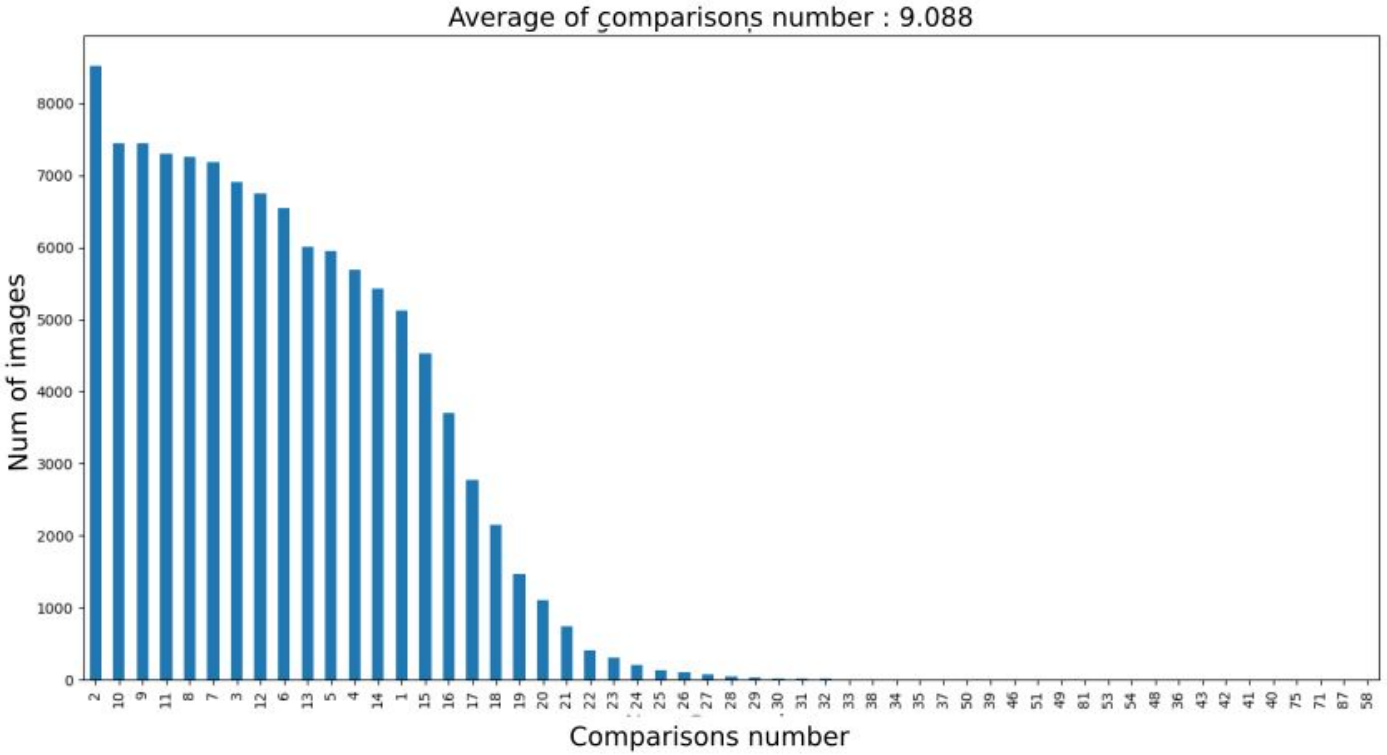
Place Pulse 2.0			
Category	# comparisons	# images	*mean*
Safety	368,926	111,389	5.188
Lively	267,292	111,348	5.085
Beautiful	175,361	110,766	4.920
Wealthy	152,241	107,795	4.890
Depressing	132,467	105,495	4.816
Boring	127,362	106,363	4.810
Total	1,223,649		

Exploratory Analysis

Number of images per continent



Number of comparisons



* **Remember:** Comparisons were made using two random images from two random cities.

Geographical city distribution: Cities included in Place Pulse 2.0

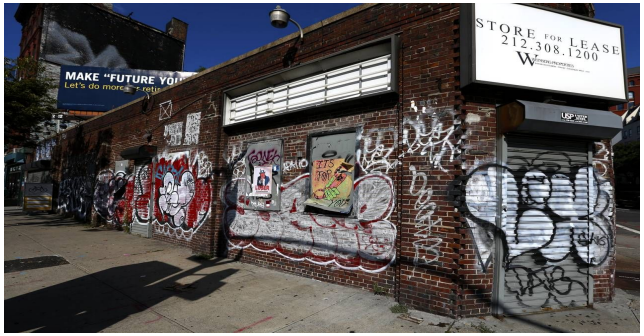
* **Note:** Same color means same country.

Number of images per geographical level

Place Pulse 2.0				
Category/Level	City	Country	Continent	Global
safety	20,143	45,640	85,890	111,390
lively	14,803	38,216	79,788	111,349
Beautiful	9,410	28,811	66,792	110,767
Wealthy	7,642	24,326	57,780	107,796
Depressing	6,556	21,171	52,504	105,496
Boring	6,148	20,931	52,031	106,364

Dataset Limitations

Individual perception

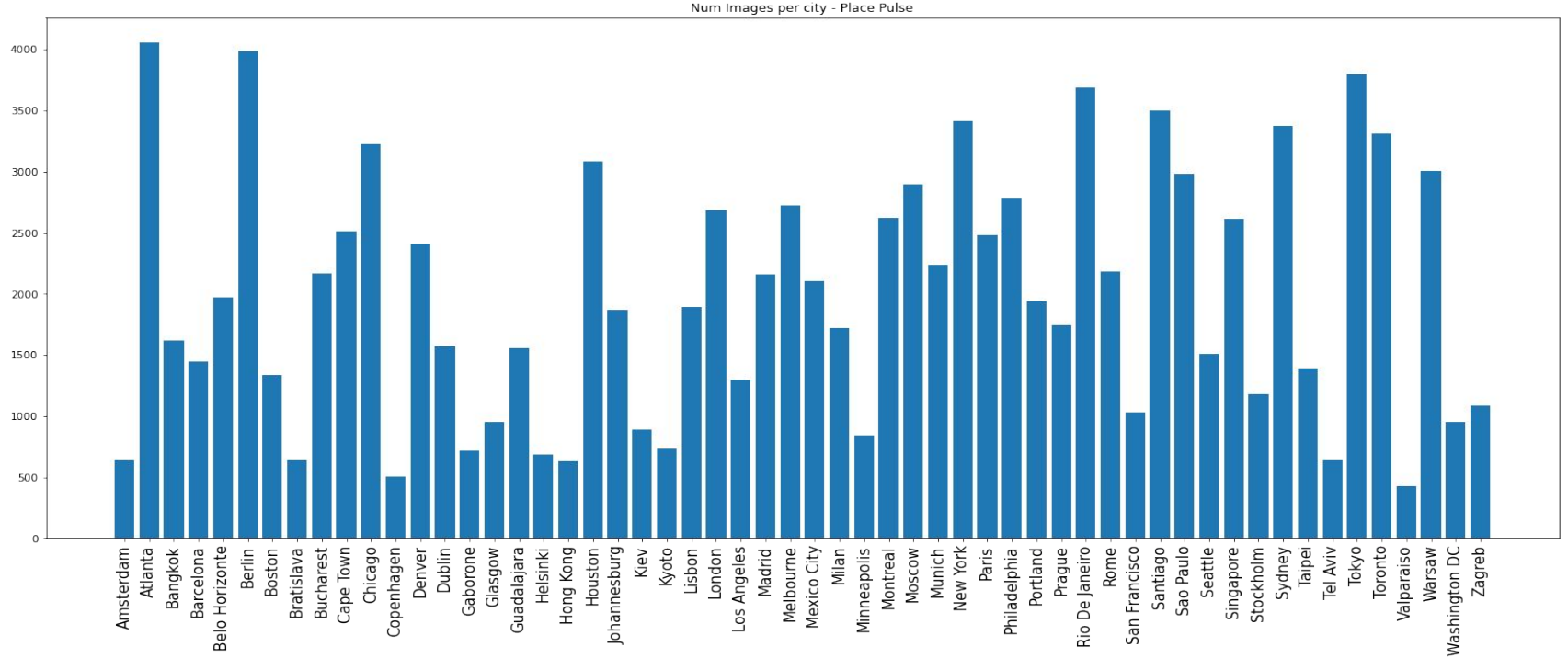
	Safe perception	Unsafe perception
New York*		
Tokyo**		

*https://www.nytimes.com/2019/08/08/nyregion/newyorktoday/times-square-panic-safety.html#:~:text=Actually%2C%20Times%20Square%20is%20one,23%2C000%20major%20crimes%20were%20recorded.

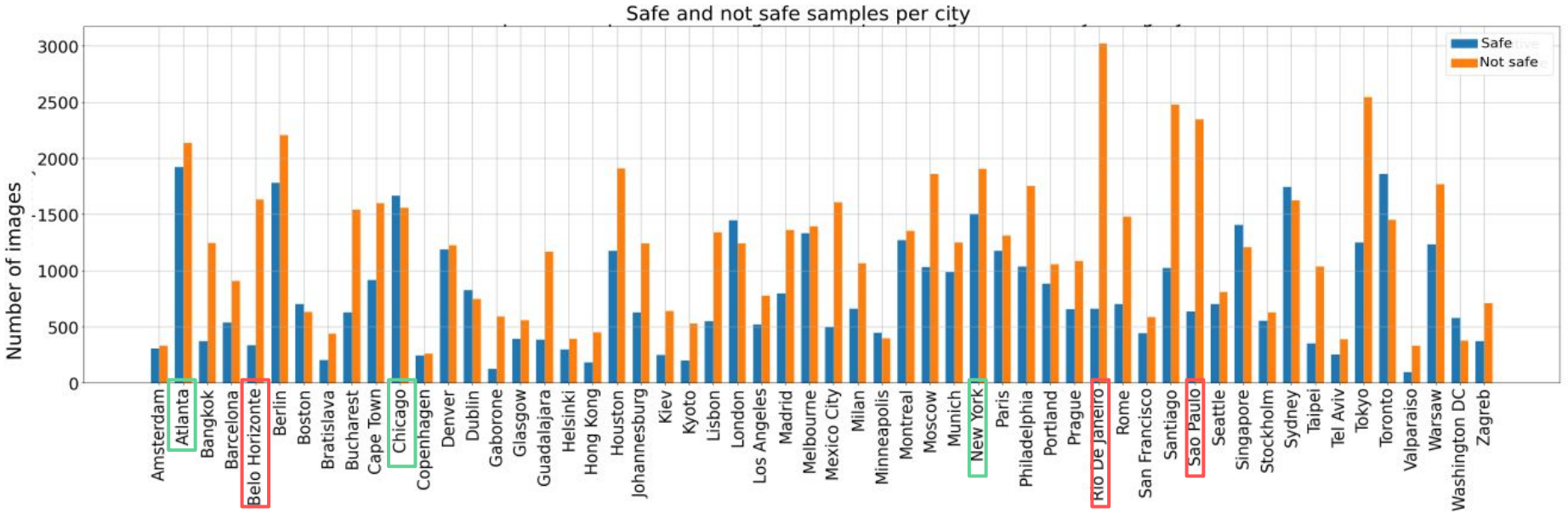
**https://www.japantimes.co.jp/news/2019/10/04/national/media-national/rip-off-bars-japan-tourist-boom/

Lack of samples: Identify city characteristics individually

Place Pulse 1.0 < 4 140 Images & Place Pulse 2.0 < 112 000 Images

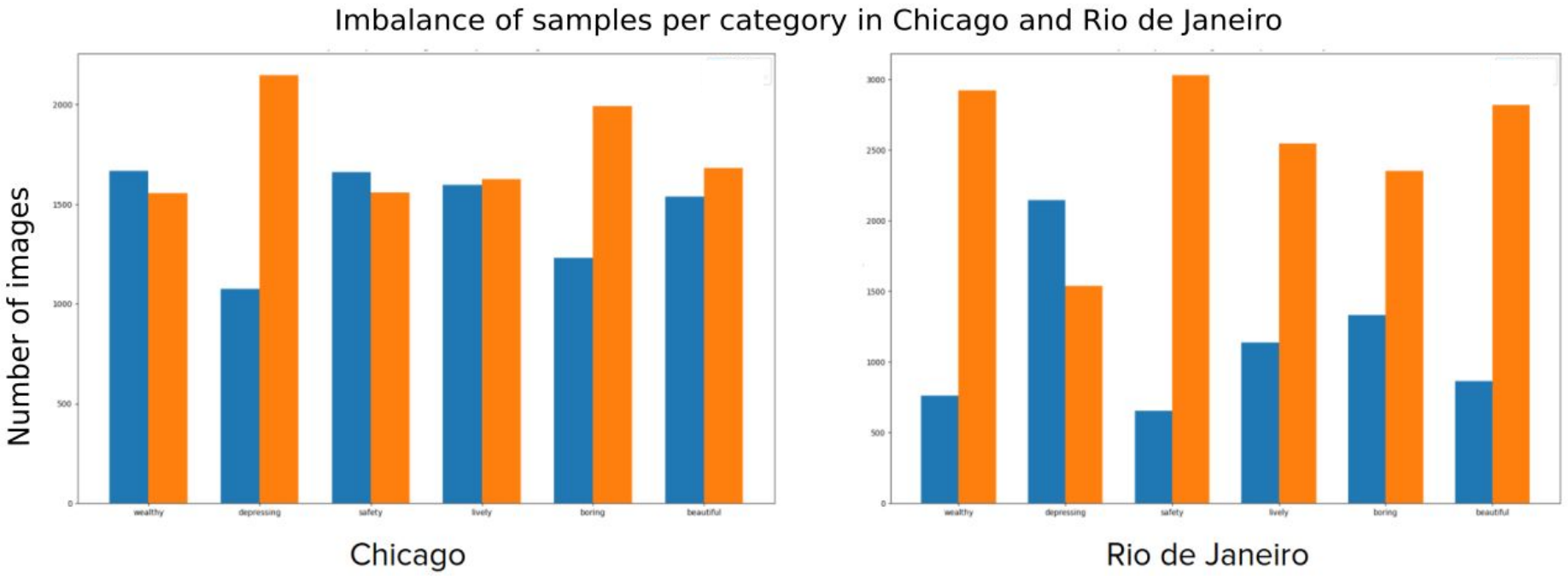


Imbalance of samples: e.g. Safety category perception



* **Note:** Some cities have more “not safe” sample than safe samples. E.g. Brazilian cities.

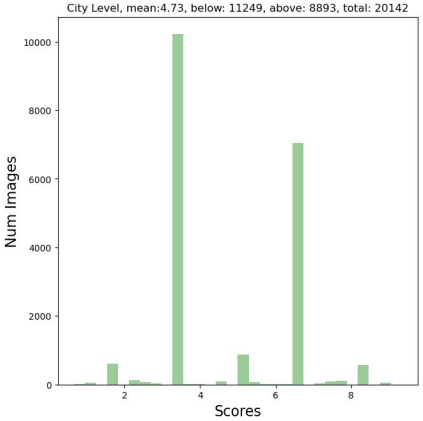
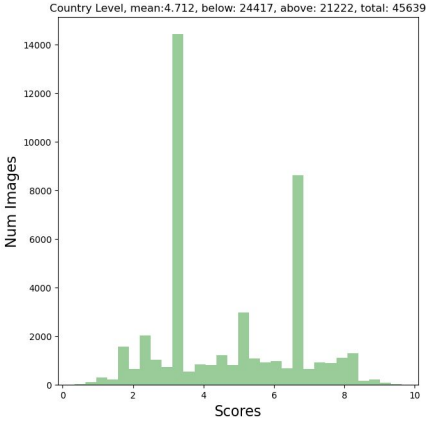
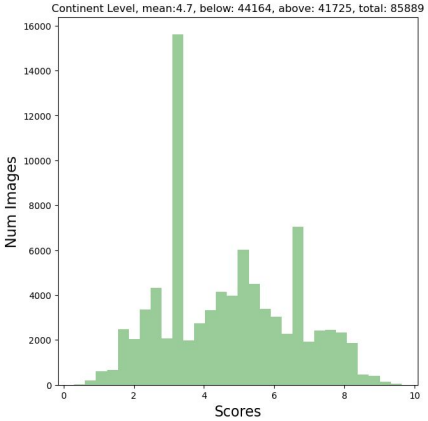
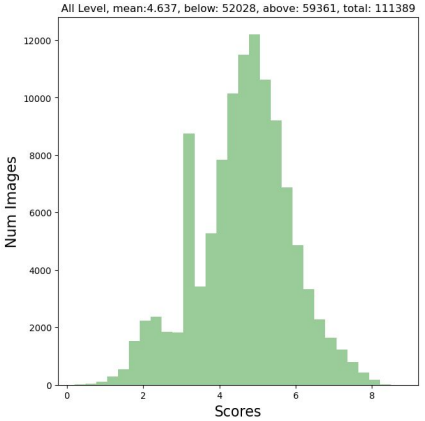
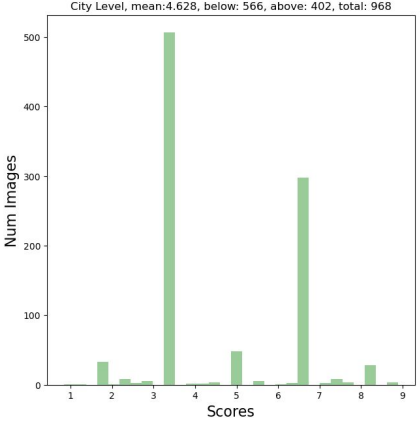
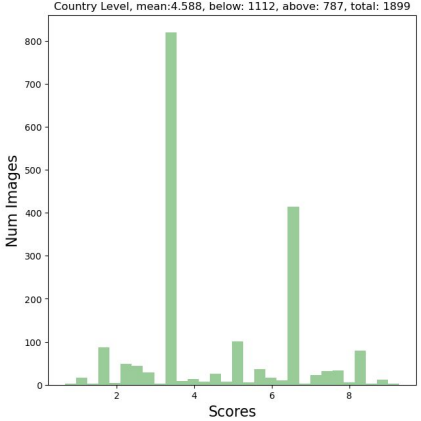
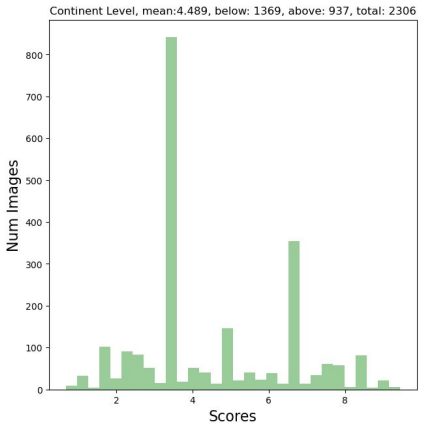
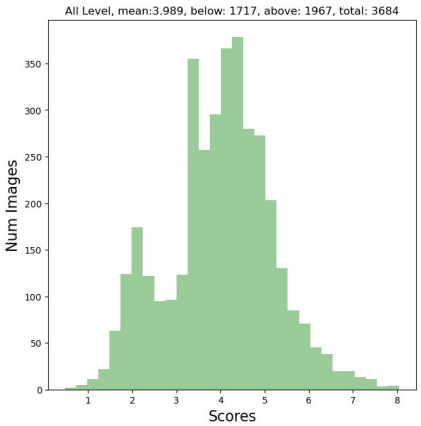
Imbalance of samples: e.g. Chicago vs Rio de Janeiro



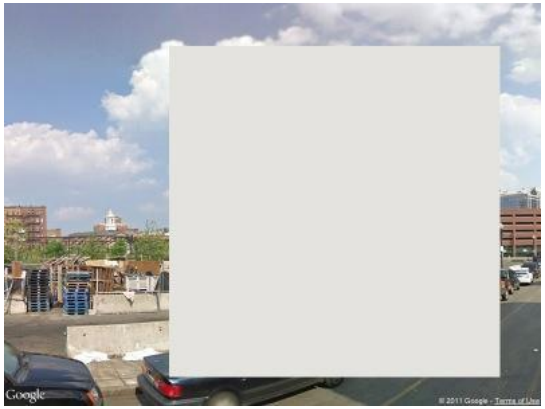
*Positive Samples: safe, beautiful, wealthy, lively, not depressing, not boring.

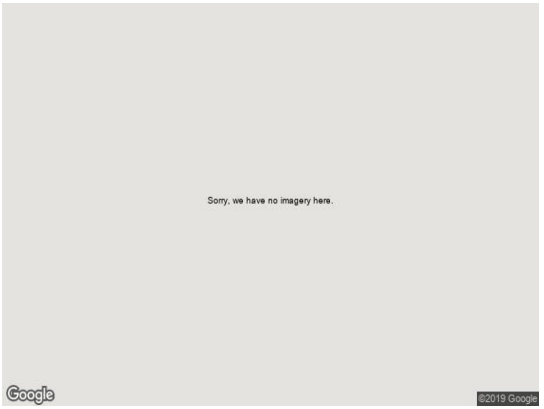
*Negative Samples: not safe, not beautiful, not wealthy, not lively, depressing, boring.

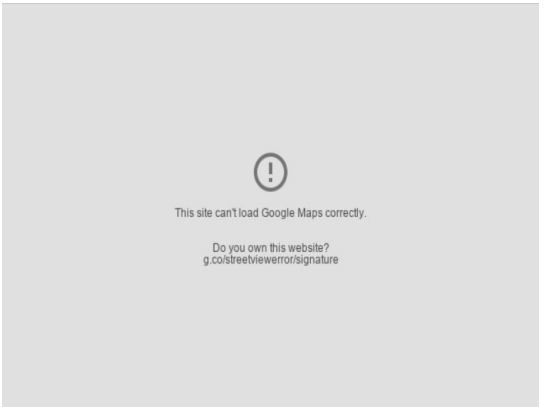
Non-Reliable Score Distribution

	City	Country	Continent	Global
World				
Rio de Janeiro				

Dataset Images: Faulty/Blank/None samples

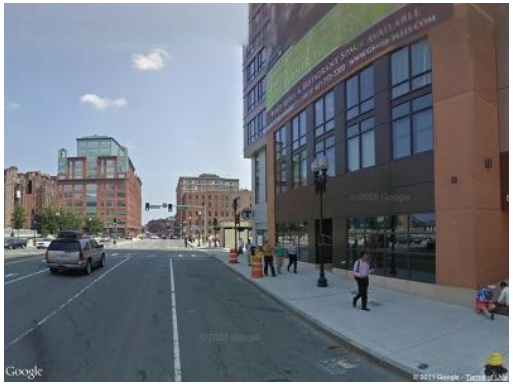






Dataset Images: Different Point of View of Sample Images



Perception changes over time

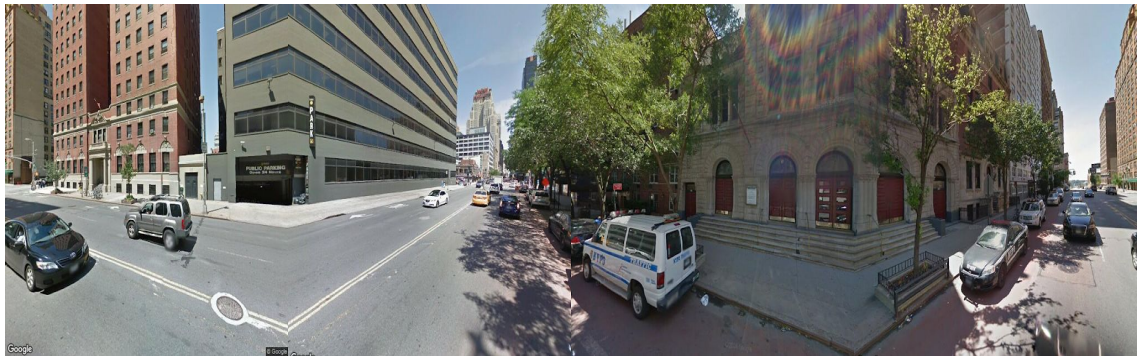
ID: 3936

ID: 1

2011 2013 2019

Different Point of View

Angle: 90



Panoramic

Experiments & Results

Metrics

- **Accuracy** — ***What percent of the data were predicted correct?***
- **Precision** — ***What percent of your predictions were correct?***
- **Recall** — ***What percent of the positive cases did you catch?***
- **F1 score** — ***What percent of positive predictions were correct?***

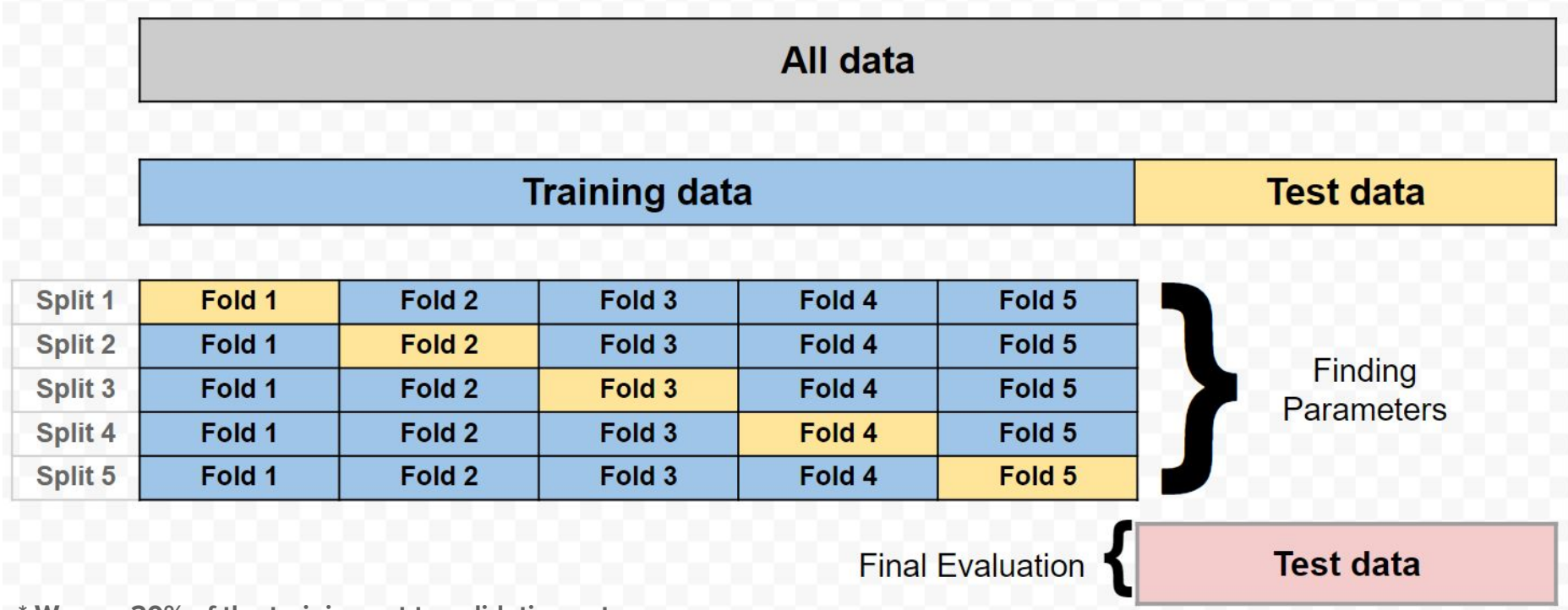
$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}$$

$$Precision = \frac{T_P}{T_P + F_P}$$

$$Recall = \frac{T_P}{T_P + F_N}$$

$$F1_{score} = 2\frac{Precision * Recall}{Precision + Recall}$$

Data Split: K-fold cross validation



* We use 20% of the training set to validation set.
* All results presented are corresponding to test data

Transfer-Learning models results

		auc		*accuracy*		*f1 score*	
Model	Method	train	eval	train	eval	entrena	eval
ResNet50	*Linear SVC*	61.62	59.10	68.10	**66.42**	53.63	50.80
	Logistic	60.04	**59.15**	67.25	66.37	51.47	49.70
	Ridge Classifier	62.11	58.38	68.36	66.08	54.59	**51.00**
	RBF SVC	45.36	44.07	53.46	53.57	44.99	44.98
Xception	*LinearSVC*	55.29	**53.25**	64.43	**63.33**	41.66	39.69
	Logistic Regression	53.48	52.75	63.56	63.14	36.72	35.87
	Ridge Classifier	57.23	52.22	65.22	63.04	45.63	42.11
	RBF SVC	45.575	44.99	49.12	49.12	55.01	**55.05**

* **Results of testing using all dataset.**

Questions?