

The background image is a wide-angle aerial photograph of the New York City skyline during a vibrant sunset. The sky is filled with large, billowing clouds that are illuminated from behind by the setting sun, creating a warm orange and yellow glow. The city's iconic skyscrapers, including the Empire State Building, the Chrysler Building, and various modern glass-fronted towers, are silhouetted against the bright sky. In the foreground, the dense grid of buildings is visible, while the Hudson River and the Manhattan Bridge are visible in the distance.

NYC AIRB&B PRICE ESTIMATION THROUGH MACHINE LEARNING

Authors: Edoardo Fazzari, Mirco Ramo

INTRODUCTION

Goals and dataset



WHY IS IMPORTANT TO ESTIMATE A CORRECT PRICE

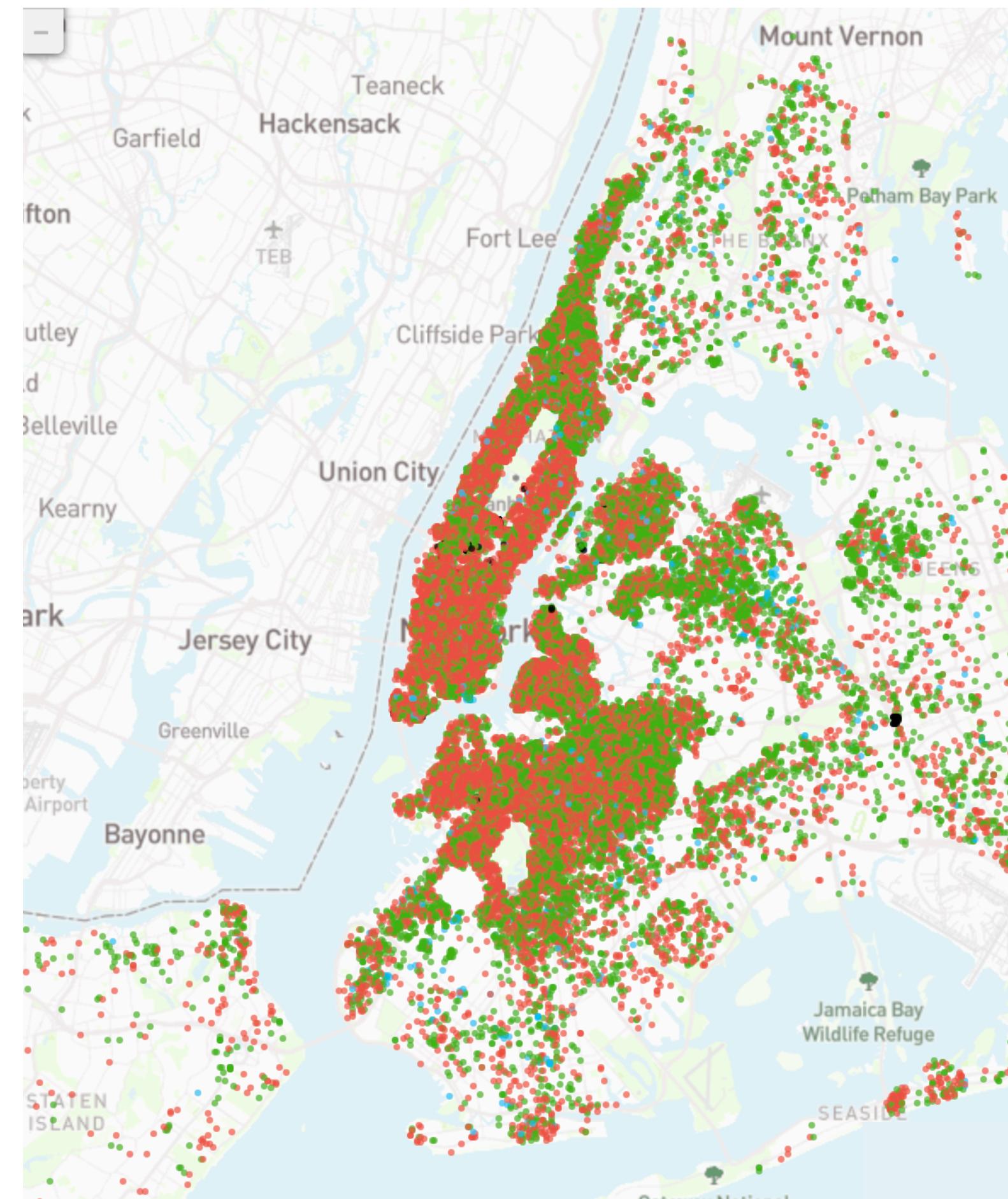
Basically for three things:

- Housing prices are important reflection of the economy
- Pricing can be challenging for the owner
- Customers have to evaluate an offered price



INITIAL DATASET

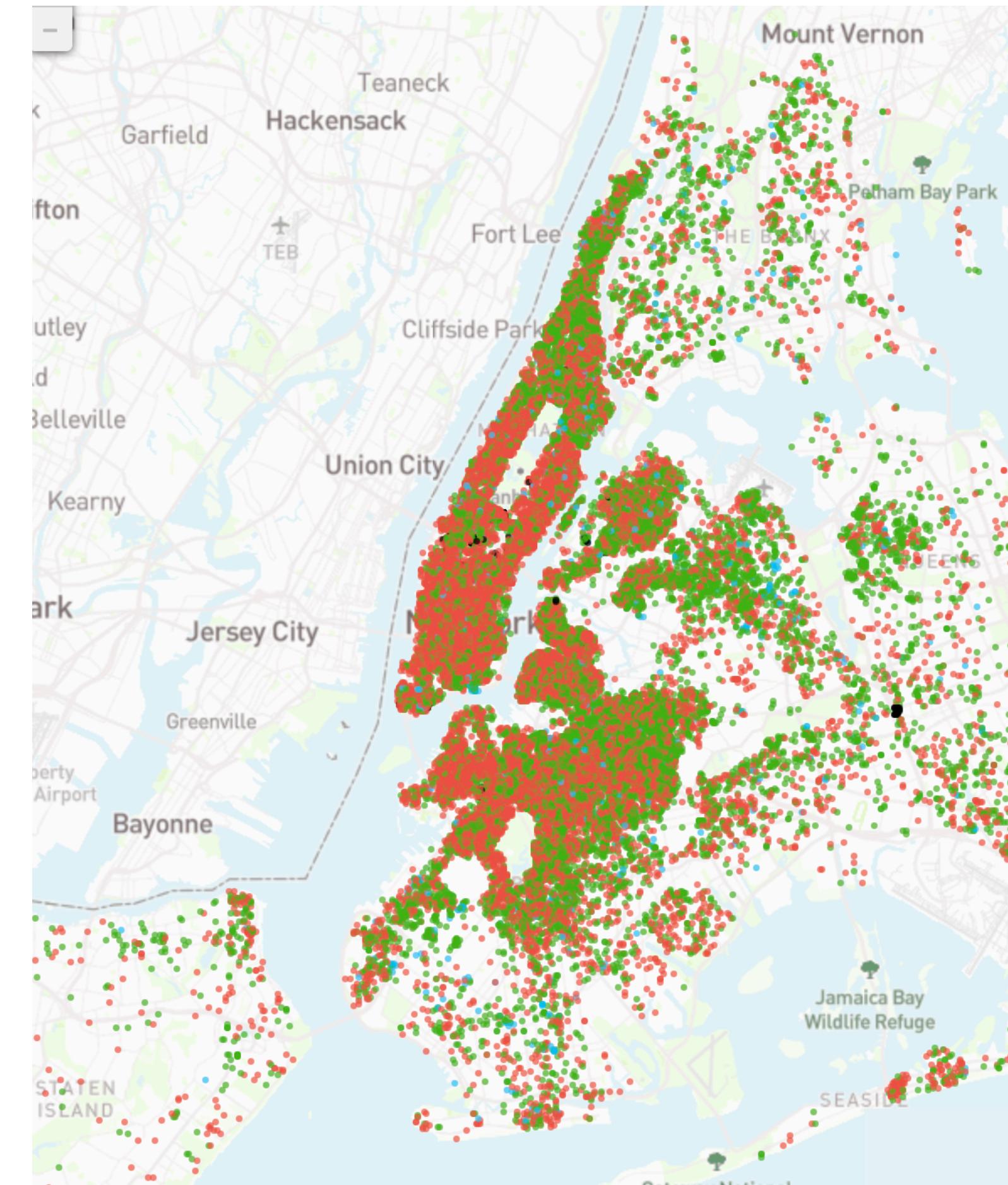
- The dataset was taken from: <http://insideairbnb.com/get-the-data.html> (November 10th 2020)
- 35821 instances and 74 columns.



INITIAL DATASET

- The dataset was taken from: <http://insideairbnb.com/get-the-data.html> (November 10th 2020)
- 35821 instances and 74 columns.

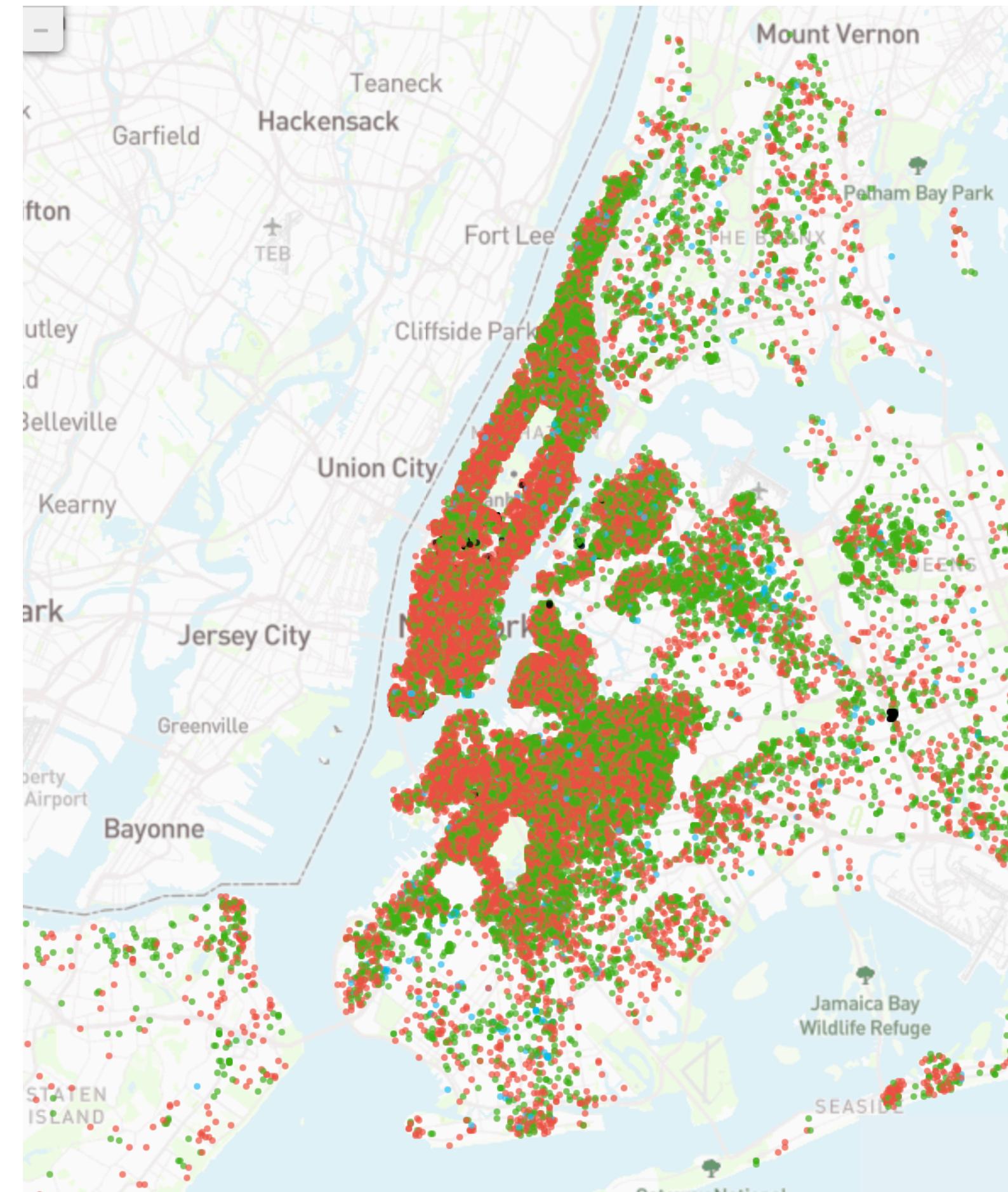
*SOME PROBLEMS RELATED TO
THE CSV FILE DOWNLOADED*



INITIAL DATASET

- The dataset was taken from: <http://insideairbnb.com/get-the-data.html> (November 10th 2020)
- 35821 instances and 74 columns.

*NEED OF
PREPROCESSING!*



PREPROCESSING

*Data Cleaning
Data Reduction
Data Transformation*



REMOVING NOISY AND IRRELEVANT ATTRIBUTES

Removed all the features that are not domain-specific

- *IDs*
- *URLs*
- *Scraping infos*
- *Useless information on rent and host*

id	listing_url	scrape_id	last_scraped	name
29800915	https://www.airbnb.com/rooms/29800915	20201210194816	2020-12-10	Entire floor (private entrance) w/ 1 BR in NYC
755528	https://www.airbnb.com/rooms/755528	20201210194816	2020-12-11	PRIVATE BATH/TONS OF SUNLIGHT/SAFE
33363084	https://www.airbnb.com/rooms/33363084	20201210194816	2020-12-11	E community that is commercially inclined.No smoke
182177	https://www.airbnb.com/rooms/182177	20201210194816	2020-12-11	A PRIVATE FLAT / APARTMENT- \$SPECIAL\$
12496118	https://www.airbnb.com/rooms/12496118	20201210194816	2020-12-11	Entire Home in NYC- 2 bedroom/ Private Entrance
17876530	https://www.airbnb.com/rooms/17876530	20201210194816	2020-12-11	Spacious Garden Apartment Private Entrance
35996222	https://www.airbnb.com/rooms/35996222	20201210194816	2020-12-11	Stylish 2 Beds Apartment with Private Parking
27343189	https://www.airbnb.com/rooms/27343189	20201210194816	2020-12-11	Jay's Studio Apartment
20904409	https://www.airbnb.com/rooms/20904409	20201210194816	2020-12-11	Luxury Tiny house • Ohka
40711239	https://www.airbnb.com/rooms/40711239	20201210194816	2020-12-11	Spacious & super clean room with Private Bathroom
24821651	https://www.airbnb.com/rooms/24821651	20201210194816	2020-12-11	Hiéroglyphe
773041	https://www.airbnb.com/rooms/773041	20201210194816	2020-12-11	Nice beautiful room In the Bronx

REMOVING REDUNDANT ATTRIBUTES

Reduction of attributes explicitly redundant (We did not compute the χ^2)

- On “neighborhood” attributes
- On “host_listings_count” attributes
- On host verification attributes
- On min/max number of nights
- On the “review” attributes

host_listings_count	host_total_listings_count
1	1
3	3
1	1
2	2
5	5
5	5
1	1
1	1
1	1
0	0

review_scores_rating	review_scores_accuracy	review_scores_cleanliness	review_scores_checkin	review_scores_communication	review_scores_location	review_scores_value
87	9	9	9	10	10	9
90	9	9	9	10	9	9
92	9	9	9	9	10	9
93	10	9	10	10	9	9
94	9	9	10	10	10	9
94	9	9	10	10	10	9
94	10	10	10	10	10	9
95	10	10	10	10	10	10
98	10	10	10	10	10	10

ADDITIONAL REMOVED FEATURES

Other columns that have been discarded are

- “Availability” attributes
- Latitude/longitude
- License (*empty*)
- Bathrooms (*empty*)
- Number of reviews within a certain amount of time

has_availability	availability_30	availability_60	availability_90
t	0	0	0
t	13	43	70
t	0	23	53
t	26	56	86
t	30	60	90
t	6	18	27
t	0	0	0

latitude	longitude	reviews_per_month
40.85948	-73.86296	
40.8584	-73.86969	
40.85791	-73.86635	
		1.79
		2.45
		5.83
		2.68

OBTAINED CSV

.....

host_since	host_response_time	host_response_rate	host_acceptance_rate	host_is_superhost	host_listings_count	host_identity_verified	neighbourhood_cleansed	property_type	room_type	accommodates	bathrooms_text	bedrooms	beds	minimum_nights	maximum_nights	number_of_reviews	review_scores_rating	instant_bookable	price
04/11/2018	?	?	100	f	1	f	Allerton	Entire_house	Entire_home-apt	4	1_bath	1	2	3	365	45	87	f	\$65.00
26/09/2012	within_an_hour	94	99	f	3	t	Allerton	Entire_apartment	Entire_home-apt	3	1_bath	1	2	2	100	242	90	f	\$49.00
09/02/2019	within_a_few_hours	57	96	f	1	t	Allerton	Entire_apartment	Entire_home-apt	3	1_bath	?	1	1	200	121	92	f	\$84.00
26/07/2011	within_a_few_hours	100	88	t	2	t	Allerton	Entire_apartment	Entire_home-apt	3	1_bath	1	2	2	28	303	93	f	\$125.00
14/01/2014	within_an_hour	80	84	f	5	t	Allerton	Entire_house	Entire_home-apt	6	1_bath	2	2	3	200	52	94	f	\$110.00
14/01/2014	within_an_hour	80	84	f	5	t	Allerton	Entire_house	Entire_home-apt	7	1_bath	2	4	4	1125	97	94	f	\$99.00
23/06/2019	within_an_hour	100	100	f	1	f	Allerton	Entire_apartment	Entire_home-apt	5	1_bath	1	2	2	15	39	94	t	\$122.00
21/07/2018	within_an_hour	100	94	t	1	f	Allerton	Entire_house	Entire_home-apt	2	1_bath	1	1	4	15	68	95	f	\$66.00
04/05/2017	within_an_hour	100	92	t	1	t	Allerton	Tiny_house	Entire_home-apt	2	1_bath	1	2	2	1125	152	98	f	\$89.00
09/12/2019	within_an_hour	100	96	f	0	t	Allerton	Private_room_in_apartment	Private_room	2	1_private_bath	1	1	1	1125	22	96	t	\$65.00
03/05/2018	within_an_hour	100	95	f	1	t	Allerton	Private_room_in_apartment	Private_room	2	1_shared_bath	1	1	30	1125	74	86	t	\$37.00
26/09/2012	within_an_hour	94	99	f	3	t	Allerton	Private_room_in_apartment	Private_room	3	1_shared_bath	1	2	30	365	214	88	t	\$38.00
17/11/2015	?	?	100	f	1	f	Allerton	Private_room_in_condominium	Private_room	1	1_shared_bath	1	1	30	1125	9	93	f	\$75.00
14/01/2014	within_an_hour	80	84	f	5	t	Allerton	Private_room_in_house	Private_room	1	1_shared_bath	1	1	30	1125	40	97	f	\$40.00
28/12/2016	?	?	?	f	1	t	Allerton	Private_room_in_house	Private_room	2	1_shared_bath	1	1	30	60	1	100	t	\$53.00
08/07/2018	?	?	?	f	0	t	Allerton	Private_room_in_house	Private_room	1	1_shared_bath	1	1	30	1125	0	?	f	\$34.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	2	1.5_baths	1	1	30	1124	68	98	t	\$55.00
23/04/2017	?	?	100	f	0	f	Allerton	Entire_house	Entire_home-apt	4	1.5_baths	2	2	30	180	1	100	t	\$145.00
23/08/2014	?	?	100	f	1	t	Allerton	Private_room_in_house	Private_room	3	1.5_shared_baths	1	1	30	90	28	91	f	\$35.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	7	1.5_shared_baths	1	4	30	1124	65	94	t	\$54.00
31/08/2017	within_an_hour	100	90	t	2	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	2	180	24	95	f	\$65.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	2	30	1124	61	97	t	\$51.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	1	1124	56	97	t	\$67.00
31/08/2017	within_an_hour	100	90	t	2	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	2	180	25	98	f	\$66.00
13/12/2012	within_an_hour	100	86	f	1	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	30	1125	1	100	f	\$78.00
13/12/2012	within_an_hour	100	86	f	1	t	Allerton	Private_room_in_house	Private_room	3	1.5_shared_baths	1	2	30	1125	0	?	f	\$68.00
14/01/2014	within_an_hour	80	84	f	5	t	Allerton	Entire_house	Entire_home-apt	12	2_baths	3	3	30	1125	25	91	f	\$205.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Entire_house	Entire_home-apt	14	2.5_baths	4	10	1	1125	90	91	t	\$249.00

DEALING WITH MISSING VALUES

Done using the Weka API:

- *Using the mean when appropriate*
- *Domain specific (e.g., bedrooms)*

OBTAINED CSV

.....

host_since	host_response_time	host_response_rate	host_acceptance_rate	host_is_superhost	host_listings_count	host_identity_verified	neighbourhood_cleansed	property_type	room_type	accommodates	bathrooms_text	bedrooms	beds	minimum_nights	maximum_nights	number_of_reviews	review_scores_rating	instant_bookable	price
04/11/2018	?	?	100	f	1	f	Allerton	Entire_house	Entire_home-apt	4	1_bath	1	2	3	365	45	87	f	\$65.00
26/09/2012	within_an_hour	94	99	f	3	t	Allerton	Entire_apartment	Entire_home-apt	3	1_bath	1	2	2	100	242	90	f	\$49.00
09/02/2019	within_a_few_hours	57	96	f	1	t	Allerton	Entire_apartment	Entire_home-apt	3	1_bath	?	1	1	200	121	92	f	\$84.00
26/07/2011	within_a_few_hours	100	88	t	2	t	Allerton	Entire_apartment	Entire_home-apt	3	1_bath	1	2	2	28	303	93	f	\$125.00
14/01/2014	within_an_hour	80	84	f	5	t	Allerton	Entire_house	Entire_home-apt	6	1_bath	2	2	3	200	52	94	f	\$110.00
14/01/2014	within_an_hour	80	84	f	5	t	Allerton	Entire_house	Entire_home-apt	7	1_bath	2	4	4	1125	97	94	f	\$99.00
23/06/2019	within_an_hour	100	100	f	1	f	Allerton	Entire_apartment	Entire_home-apt	5	1_bath	1	2	2	15	39	94	t	\$122.00
21/07/2018	within_an_hour	100	94	t	1	f	Allerton	Entire_house	Entire_home-apt	2	1_bath	1	1	4	15	68	95	f	\$66.00
04/05/2017	within_an_hour	100	92	t	1	t	Allerton	Tiny_house	Entire_home-apt	2	1_bath	1	2	2	1125	152	98	f	\$89.00
09/12/2019	within_an_hour	100	96	f	0	t	Allerton	Private_room_in_apartment	Private_room	2	1_private_bath	1	1	1	1125	22	96	t	\$65.00
03/05/2018	within_an_hour	100	95	f	1	t	Allerton	Private_room_in_apartment	Private_room	2	1_shared_bath	1	1	30	1125	74	86	t	\$37.00
26/09/2012	within_an_hour	94	99	f	3	t	Allerton	Private_room_in_apartment	Private_room	3	1_shared_bath	1	2	30	365	214	88	t	\$38.00
17/11/2015	?	?	100	f	1	f	Allerton	Private_room_in_condominium	Private_room	1	1_shared_bath	1	1	30	1125	9	93	f	\$75.00
14/01/2014	within_an_hour	80	84	f	5	t	Allerton	Private_room_in_house	Private_room	1	1_shared_bath	1	1	30	1125	40	97	f	\$40.00
28/12/2016	?	?	?	f	1	t	Allerton	Private_room_in_house	Private_room	2	1_shared_bath	1	1	30	60	1	100	t	\$53.00
08/07/2018	?	?	?	f	0	t	Allerton	Private_room_in_house	Private_room	1	1_shared_bath	1	1	30	1125	0	?	f	\$34.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	2	1.5_baths	1	1	30	1124	68	98	t	\$55.00
23/04/2017	?	?	100	f	0	f	Allerton	Entire_house	Entire_home-apt	4	1.5_baths	2	2	30	180	1	100	t	\$145.00
23/08/2014	?	?	100	f	1	t	Allerton	Private_room_in_house	Private_room	3	1.5_shared_baths	1	1	30	90	28	91	f	\$35.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	7	1.5_shared_baths	1	4	30	1124	65	94	t	\$54.00
31/08/2017	within_an_hour	100	90	t	2	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	2	180	24	95	f	\$65.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	2	30	1124	61	97	t	\$51.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	1	1124	56	97	t	\$67.00
31/08/2017	within_an_hour	100	90	t	2	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	2	180	25	98	f	\$66.00
13/12/2012	within_an_hour	100	86	f	1	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	30	1125	1	100	f	\$78.00
13/12/2012	within_an_hour	100	86	f	1	t	Allerton	Private_room_in_house	Private_room	3	1.5_shared_baths	1	2	30	1125	0	?	f	\$68.00
14/01/2014	within_an_hour	80	84	f	5	t	Allerton	Entire_house	Entire_home-apt	12	2_baths	3	3	30	1125	25	91	f	\$205.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Entire_house	Entire_home-apt	14	2.5_baths	4	10	1	1125	90	91	t	\$249.00

OBTAINED CSV

.....

host_since	host_response_time	host_response_rate	host_acceptance_rate	host_is_superhost	host_listings_count	host_identity_verified	neighbourhood_cleansed	property_type	room_type	accommodates	bathrooms_text	bedrooms	beds	minimum_nights	maximum_nights	number_of_reviews	review_scores_rating	instant_bookable	price
04/11/2018	?	?	100	f	1	f	Allerton	Entire_house	Entire_home-apt	4	1_bath	1	2	3	365	45	87	f	\$65.00
26/09/2012	within_an_hour	94	99	f	3	t	Allerton	Entire_apartment	Entire_home-apt	3	1_bath	1	2	2	100	242	90	f	\$49.00
09/02/2019	within_a_few_hours	57	96	f	1	t	Allerton	Entire_apartment	Entire_home-apt	3	1_bath	?	1	1	200	121	92	f	\$84.00
26/07/2011	within_a_few_hours	100	88	t	2	t	Allerton	Entire_apartment	Entire_home-apt	3	1_bath	1	2	2	28	303	93	f	\$125.00
14/01/2014	within_an_hour	80	84	f	5	t	Allerton	Entire_house	Entire_home-apt	6	1_bath	2	2	3	200	52	94	f	\$110.00
14/01/2014	within_an_hour	80	84	f	5	t	Allerton	Entire_house	Entire_home-apt	7	1_bath	2	4	4	1125	97	94	f	\$99.00
23/06/2019	within_an_hour	100	100	f	1	f	Allerton	Entire_apartment	Entire_home-apt	5	1_bath	1	2	2	15	39	94	t	\$122.00
21/07/2018	within_an_hour	100	94	t	1	f	Allerton	Entire_house	Entire_home-apt	2	1_bath	1	1	4	15	68	95	f	\$66.00
04/05/2017	within_an_hour	100	92	t	1	t	Allerton	Tiny_house	Entire_home-apt	2	1_bath	1	2	2	1125	152	98	f	\$89.00
09/12/2019	within_an_hour	100	96	f	9	t	Allerton	Private_room_in_apartment	Private_room	9	1_private_bath	1	1	1	1125	22	96	t	\$65.00
03/05/2018	within_an_hour	100													1125	74	86	t	\$37.00
26/09/2012	within_an_hour	94													365	214	88	t	\$38.00
17/11/2015	?	?													1125	9	93	f	\$75.00
14/01/2014	within_an_hour	80													1125	40	97	f	\$40.00
28/12/2016	?	?	?												60	1	100	t	\$53.00
08/07/2018	?	?	?												1125	0	?	f	\$34.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	2	1.5_baths	1	1	30	1124	68	98	t	\$55.00
23/04/2017	?	?	100	f	0	f	Allerton	Entire_house	Entire_home-apt	4	1.5_baths	2	2	30	180	1	100	t	\$145.00
23/08/2014	?	?	100	f	1	t	Allerton	Private_room_in_house	Private_room	3	1.5_shared_baths	1	1	30	90	28	91	f	\$35.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	7	1.5_shared_baths	1	4	30	1124	65	94	t	\$54.00
31/08/2017	within_an_hour	100	90	t	2	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	2	180	24	95	f	\$65.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	2	30	1124	61	97	t	\$51.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	1	1124	56	97	t	\$67.00
31/08/2017	within_an_hour	100	90	t	2	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	2	180	25	98	f	\$66.00
13/12/2012	within_an_hour	100	86	f	1	t	Allerton	Private_room_in_house	Private_room	2	1.5_shared_baths	1	1	30	1125	1	100	f	\$78.00
13/12/2012	within_an_hour	100	86	f	1	t	Allerton	Private_room_in_house	Private_room	3	1.5_shared_baths	1	2	30	1125	0	?	f	\$68.00
14/01/2014	within_an_hour	80	84	f	5	t	Allerton	Entire_house	Entire_home-apt	12	2_baths	3	3	30	1125	25	91	f	\$205.00
24/05/2018	within_an_hour	100	100	f	5	t	Allerton	Entire_house	Entire_home-apt	14	2.5_baths	4	10	1	1125	90	91	t	\$249.00

WHERE ARE THE AMENITIES?

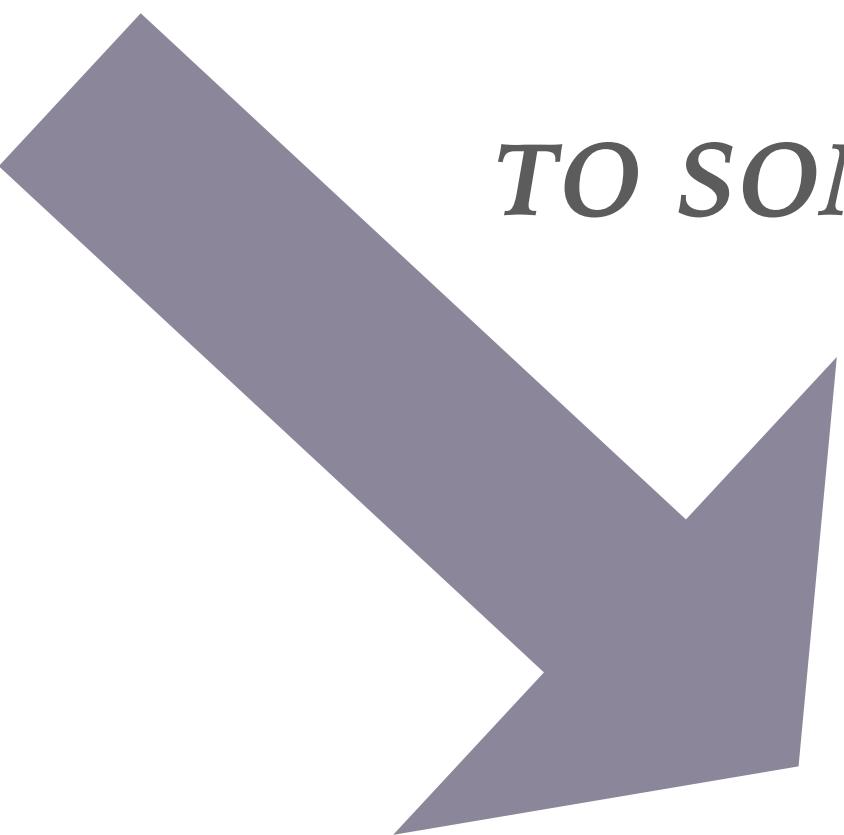
DATA TRANSFORMATION: AMENITIES

amenities
["Hot water", "Shampoo", "Oven", "Single level home", "Iron", "Long term stays allowed", "Wifi", "Refrigerator", "Dedicated workspace", "Air conditioning", "Free street parking", "Smoke alarm"]
["Hot water", "Shampoo", "Iron", "TV", "Wifi", "Refrigerator", "Dedicated workspace", "Air conditioning", "Keypad", "Free street parking", "Smoke alarm", "Heating", "Fire extinguisher", "Essentials"]
["Dishes and silverware", "Heating", "Shampoo", "Essentials", "Air conditioning", "Hair dryer", "Hangers", "Iron", "Free street parking", "Kitchen", "TV", "Carbon monoxide alarm", "Smoke alarm"]
["Hot water", "Shampoo", "Game console", "Oven", "Extra pillows and blankets", "Single level home", "Laundromat nearby", "Shower gel", "Iron", "TV", "Wifi", "Refrigerator", "Dedicated workspace", "Air conditioning", "Free street parking", "Lockbox"]
["Hot water", "Shampoo", "Extra pillows and blankets", "Oven", "Single level home", "Iron", "TV", "Baking sheet", "Wifi", "Refrigerator", "Dedicated workspace", "Air conditioning", "Free street parking", "Lockbox"]
["Hot water", "Shampoo", "Oven", "Iron", "Long term stays allowed", "TV", "Wifi", "Refrigerator", "Dedicated workspace", "Air conditioning", "Keypad", "Free street parking", "Smoke alarm", "Dishwasher", "Heating", "Parking on premises"]
["Hot water", "Shampoo", "Oven", "Iron", "TV", "Wifi", "Refrigerator", "Air conditioning", "Free parking on premises", "Free street parking", "Smoke alarm", "Dishes and silverware", "Heating", "Parking on premises"]
["Hot water", "Shampoo", "Extra pillows and blankets", "Oven", "Single level home", "Iron", "Washer", "TV", "Wifi", "Refrigerator", "Dryer", "Dedicated workspace", "Air conditioning", "Lockbox"]
["Hot water", "Shampoo", "Single level home", "Iron", "Long term stays allowed", "TV", "Wifi", "Refrigerator", "Dedicated workspace", "Air conditioning", "Free street parking", "Smoke alarm", "Parking on premises"]
["Hot water", "Shampoo", "Elevator", "Iron", "Long term stays allowed", "TV", "Wifi", "Refrigerator", "Dedicated workspace", "Air conditioning", "Lock on bedroom door", "Free street parking", "Parking on premises"]
["Hot water", "Oven", "Shower gel", "Iron", "Long term stays allowed", "TV", "Wifi", "Refrigerator", "Air conditioning", "Lock on bedroom door", "Free street parking", "Smoke alarm", "Bathtub", "Parking on premises"]
["Hot water", "Shampoo", "Oven", "Iron", "TV", "Wifi", "Refrigerator", "Dedicated workspace", "Air conditioning", "Lock on bedroom door", "Free street parking", "Smoke alarm", "First aid kit", "Parking on premises"]
["Hot water", "Shampoo", "Extra pillows and blankets", "Oven", "Elevator", "Iron", "Washer", "TV", "Wifi", "Refrigerator", "Dryer", "Dedicated workspace", "Dishwasher", "Free street parking", "Parking on premises"]
["Wifi", "Heating", "Hot water", "Shampoo", "Essentials", "Cable TV", "Dedicated workspace", "Air conditioning", "Hair dryer", "Lock on bedroom door", "Hangers", "Keypad", "Iron", "Free street parking", "Parking on premises"]
["Wifi", "Heating", "Private living room", "Dedicated workspace", "Air conditioning", "Free parking on premises", "Indoor fireplace", "Hangers", "Iron", "Kitchen", "TV"]
["Wifi", "Heating", "Shampoo", "Essentials", "Dedicated workspace", "Air conditioning", "Hangers", "Iron", "Kitchen", "TV"]
["Hot water", "Shampoo", "Oven", "Patio or balcony", "Paid parking on premises", "Iron", "Washer", "Long term stays allowed", "TV", "Wifi", "Refrigerator", "Dryer", "Air conditioning", "Dedicated workspace", "Parking on premises"]
["Hot water", "Shampoo", "Oven", "Patio or balcony", "Shower gel", "Iron", "TV", "Wifi", "Refrigerator", "Air conditioning", "Free street parking", "Smoke alarm", "Dishes and silverware", "Heating", "Parking on premises"]
["Hot water", "Oven", "Patio or balcony", "Cleaning before checkout", "Iron", "Long term stays allowed", "TV", "Wifi", "Refrigerator", "Dedicated workspace", "Air conditioning", "Lockbox", "Lock on bedroom door", "Parking on premises"]
["Wifi", "Heating", "Hot water", "Shampoo", "Essentials", "Dryer", "Paid parking on premises", "Dedicated workspace", "Air conditioning", "Free street parking", "Hair dryer", "Breakfast", "Lockbox", "Lock on bedroom door", "Parking on premises"]
["Hot water", "Shampoo", "Shower gel", "Iron", "Wifi", "Refrigerator", "Dedicated workspace", "Air conditioning", "Lockbox", "Lock on bedroom door", "Free street parking", "Smoke alarm", "First aid kit", "Parking on premises"]

ONE HOT FOR THE AMENITIES

["Wifi", "Air conditioning", "Kitchen", "Cable TV", "TV", "Elevator", "Heating"]

["Cable TV", "Essentials", "Washer", "Heating", "Air conditioning", "Kitchen", "TV", "Dryer", "Wifi"]



TO SOMETHING LIKE THIS

Cleaning_before_checkout	Coffee_machine	Concierge	Crib
1	1	0	0
0	1	0	1
0	0	0	0
0	0	0	0

ONE HOT REFINEMENTS

Irrelevant columns:

- Limited housekeeping $\cup 2014$ on request
- Safe

Doss Bluetooth Speaker	Bluetooth sound system	Dove body soap	Dr Woods body soap
	0	0	0
	0	0	0
	0	0	0

Columns that were similar:

- Merged together

$$\text{new column} = \sum_i^{\text{similar columns}} \text{columns}_i$$

Goldies Natural Beauty shampoo	Golez conditioner	Google Home Bluetooth sound system
0	0	0
0	0	0
0	0	0
0	0	0

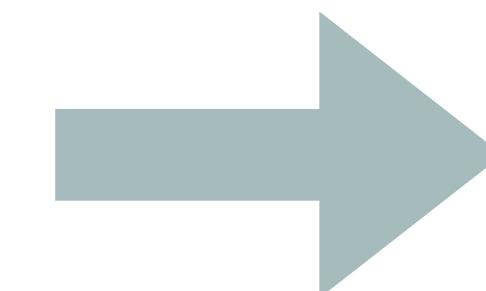
DATA TRANSFORMATION: BATHROOMS & PRICE

Bathrooms:

- Only in text format
 - Separate the information in two columns: “the # of bathrooms” and “shared”

Price:

- *Removed the dollar sign*



CLASSIFICATION

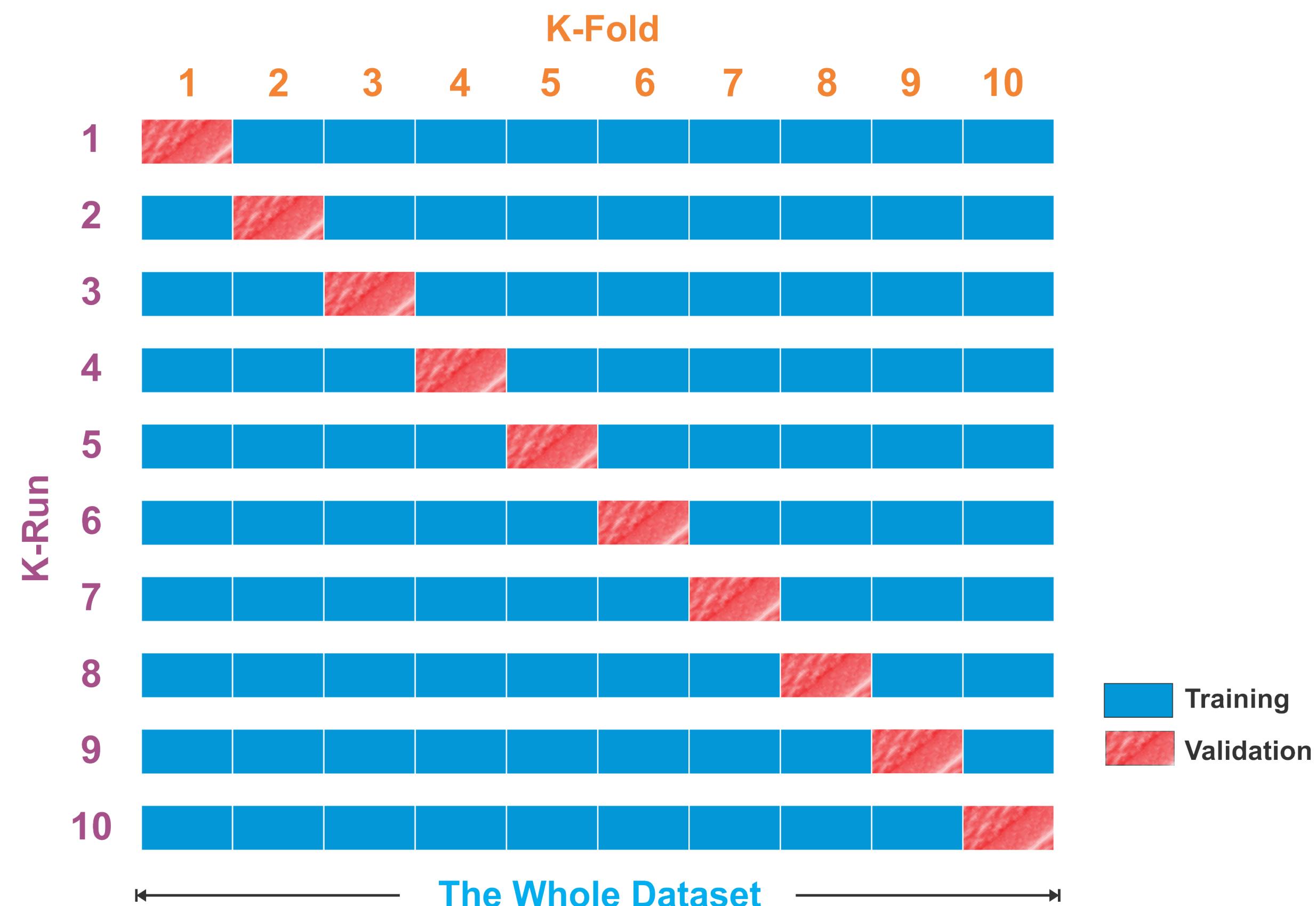
How we built our classifiers



TRAIN AND TEST SPLITTING

All the classifiers have been evaluated with the same strategy

- *10-fold cross validation*
 - *90% training set*
 - *10% test set*



TRAIN AND TEST SPLITTING

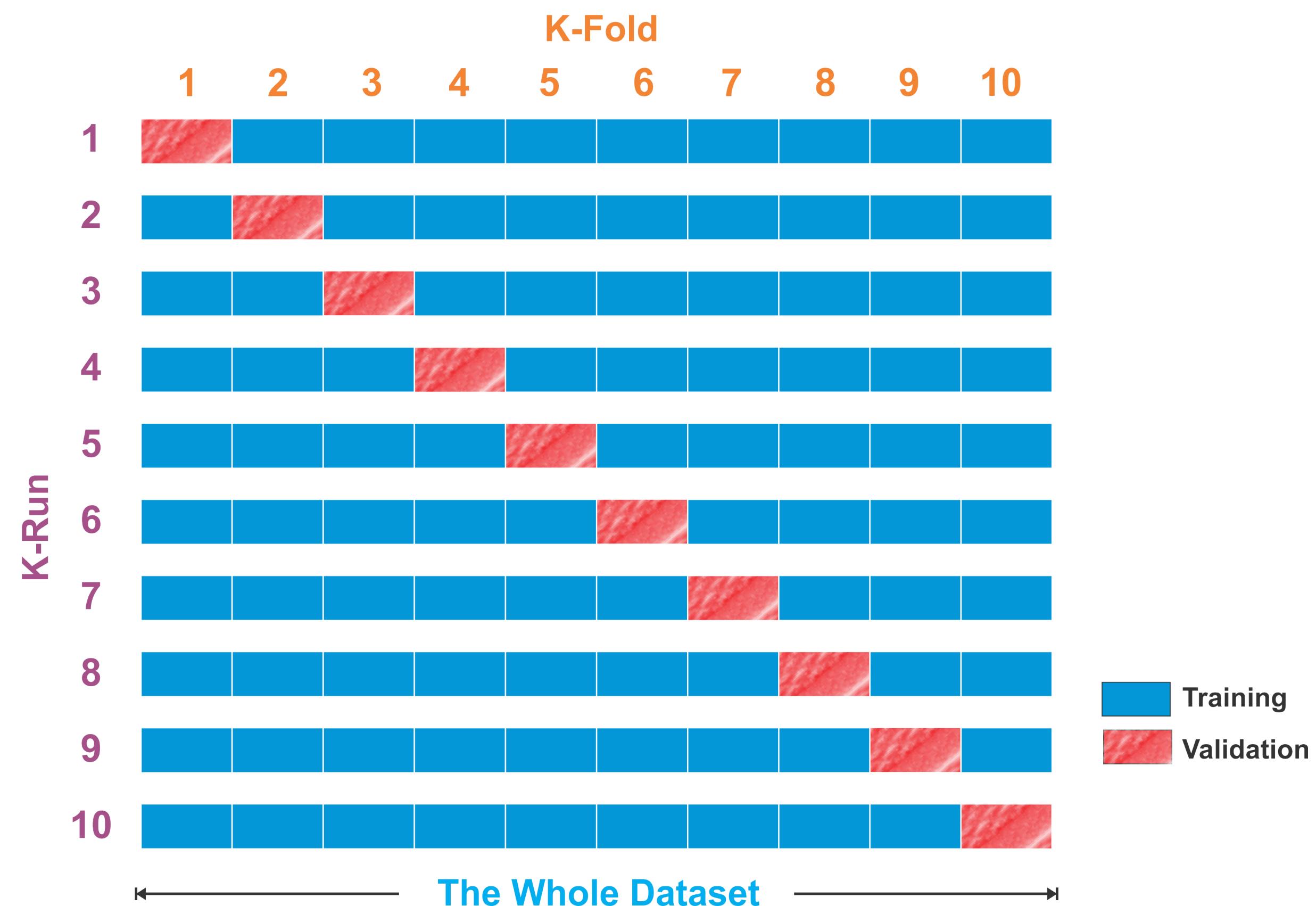
All the classifiers have been evaluated with the same strategy

- *10-fold cross validation*

- *90% training set*

- *10% test set*

We wanted to save all the result of every fold



TRAIN AND TEST SPLITTING

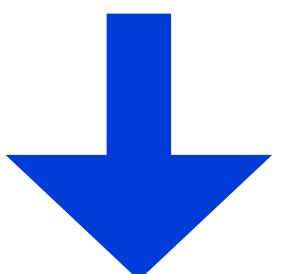
All the classifiers have been evaluated with the same strategy

➤ *10-fold cross validation*

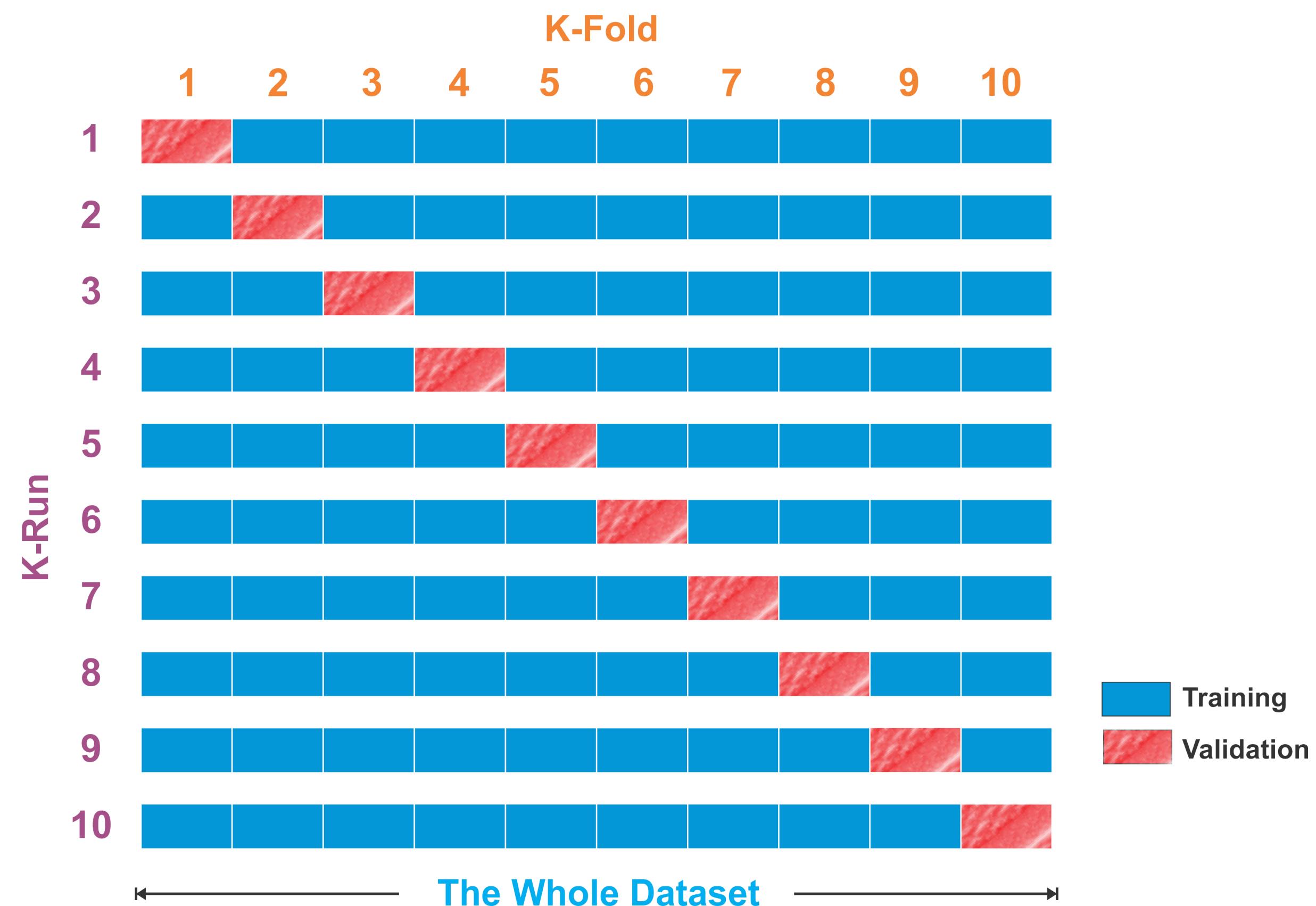
→ *90% training set*

→ *10% test set*

We wanted to save all the result of every fold



Evaluation.crossValidateModel()



TRAIN AND TEST SPLITTING

All the classifiers have been evaluated with the same strategy

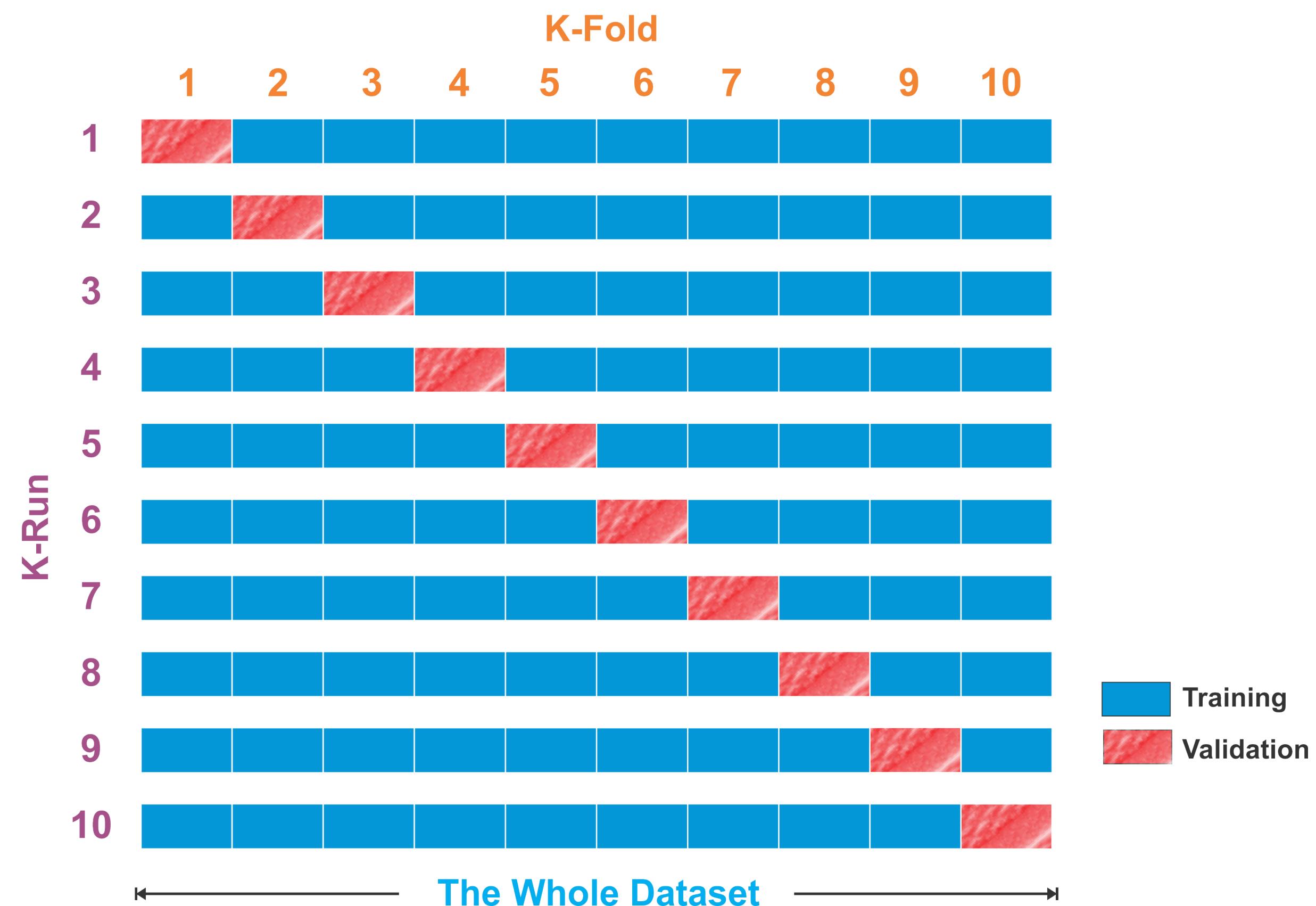
- *10-fold cross validation*

- *90% training set*

- *10% test set*

We wanted to save all the result of every fold

 Evaluation.createValidateModel()



TRAIN AND TEST SPLITTING

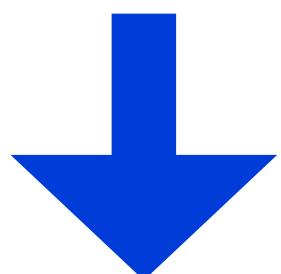
All the classifiers have been evaluated with the same strategy

- *10-fold cross validation*

- *90% training set*

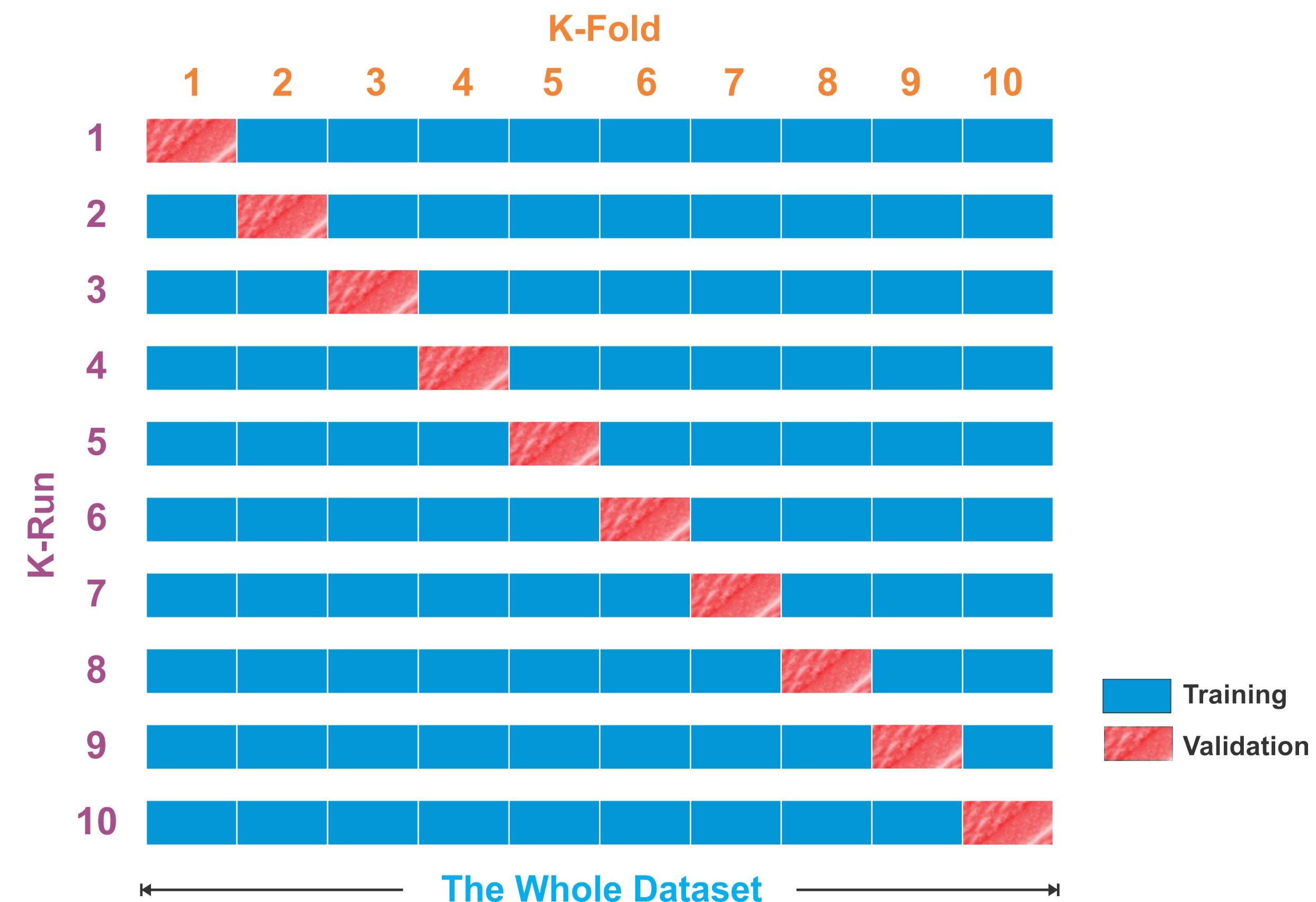
- *10% test set*

We wanted to save all the result of every fold



`Instances.trainCV(numFolds, currentFold)`

`Instances.testCV(numFolds, currentFold)`



CHOSEN CLASSIFIERS

The classifiers needed are actually regression algorithms

- *Tested on*
 - ➔ *Full dataset (138 features)*
 - ➔ *Reduced subspaces of the training set*
 - *CfsSubsetEval + BestFirst*
 - *CfsSubsetEval + GreedStepwise*

Why not InfoGain and PCA?

CHOSEN CLASSIFIERS

Classifiers tested:

- *Linear Regression w/o attribute selection*
- *Random Forest w/o attribute selection*
- *5-NN w/o attribute selection*
- *M5Rules w/o attribute selection*

PERFORMANCE EVALUATION AND AFFECT OF ATTRIBUTE SELECTION

ATTRIBUTE SELECTION REGRESSION ALGORITHM	→ ↓	NONE	CfsSubsetEval + BestFirst	CfsSubsetEval + GreedyStepwise
Linear Regression		Out of memory	CC = 0.6793 MAE = 40.0323 RMSE = 57,8399 RAE = 66.9399 RRSE = 73.3966	CC = 0.6794 MAE = 40.0308 RMSE = 57,8339 RAE = 66.9375 RRSE = 73.3891
Random Forest			CC = 0.7495 MAE = 36.0881 RMSE = 52.8727 RAE = 60.3423 RRSE = 67.0863	CC = 0.6988 MAE = 37.9716 RMSE = 56.3332 RAE = 63.4916 RRSE = 71.7302
5-KK			CC = 0.5413 MAE = 46.4003 RMSE = 67.0074 RAE = 77.2522 RRSE = 84.3622	CC = 0.6772 MAE = 39.0709 RMSE = 58.0759 RAE = 65.3630 RRSE = 73.6845
M5Rules			CC = 0.6589 MAE = 567.0975 RMSE = 31'491.7374 RAE = 942.2971 RRSE = 39'372.3750	CC= 0.6942 MAE = 38.5873 RMSE = 56.7530 RAE = 64.5234 RRSE = 72.0102

EVALUATION OF SIGNIFICANCE OF THE CLASSIFIERS' RESULTS

Student's t-test on RMSE (only CfsSubsetEval + GreedyStepwise)

- *Random Forest/Linear Regression*

$$t = 5.5342 \quad p = 0.004$$

- *Random Forest/M5Rules*

$$t = 0.805 \quad p = 0.4413$$

- *M5Rules/Linear Regression*

$$t = 5.5104 \quad p = 0.032$$

- *Linear Regression/KNN*

$$t = 0.6601 \quad p = 0.5257$$

EVALUATION OF SIGNIFICANCE OF THE CLASSIFIERS' RESULTS

Student's t-test on RMSE (only CfsSubsetEval + GreedyStepwise)

➤ *Random Forest/Linear Regression*

$$t = 5.5342$$

$$p = 0.004$$



stat. sign.

➤ *Random Forest/M5Rules*

$$t = 0.805$$

$$p = 0.4413$$



Non stat. sign.

➤ *M5Rules/Linear Regression*

$$t = 5.5104$$

$$p = 0.032$$



stat. sign.

➤ *Linear Regression/KNN*

$$t = 0.6601$$

$$p = 0.5257$$



Non stat. sign.