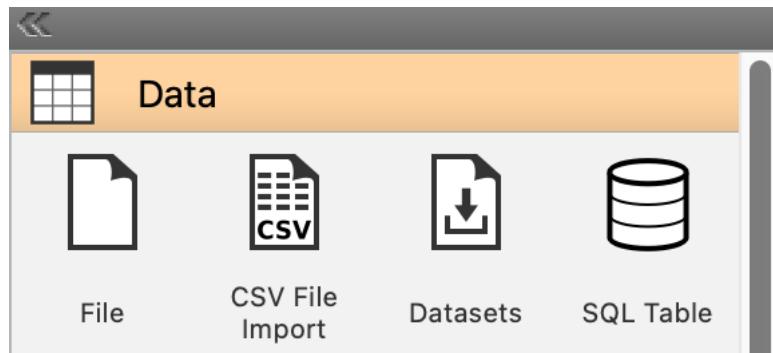


Langkah pertama yang perlu dilakukan adalah dengan melakukan input data ke Orange terlebih dahulu melalui fitur File yang dapat dilihat pada gambar berikut.



Gambar 2. 1 Fitur File

Langkah selanjutnya, lakukan klik pada fitur File, sehingga pada lembar kerja Orange akan muncul secara otomatis ikon File seperti pada gambar berikut.



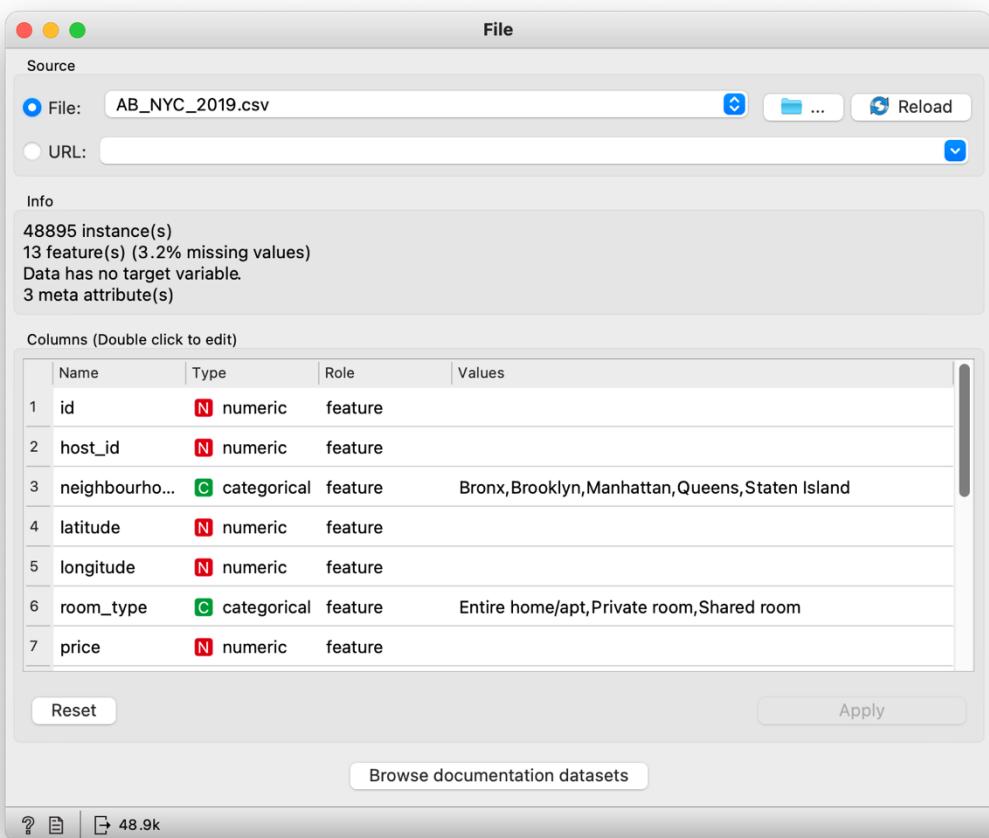
Gambar 2. 2 ikon Fitur File

Selanjutnya, lakukan klik ganda pada ikon Folder pada tampilan seperti pada gambar berikut dan pilih data AB\_NYC\_2019.csv untuk data AirBnB.



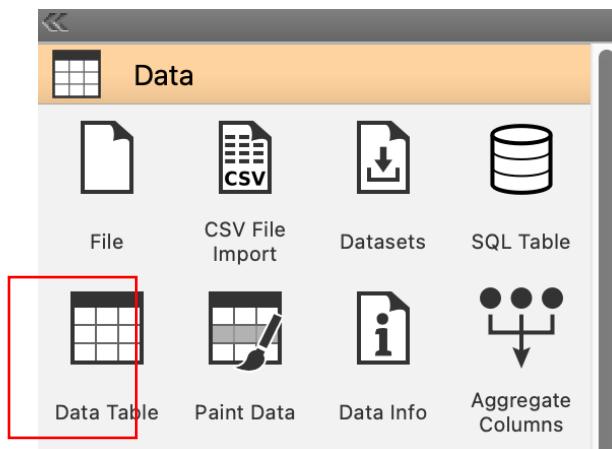
Gambar 2. 3 Klik ikon folder

Sehingga nama Source File yang tampil sudah sesuai dengan data yang akan dilakukan analisis seperti pada gambar berikut.



Gambar 2. 4 Tampilan akhir pada ikon fitur File

Langkah selanjutnya, klik fitur Data Table seperti pada gambar berikut.



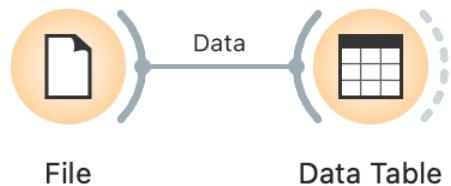
Gambar 2. 5 Fitur Data Table

Sehingga pada lembar kerja Orange akan terlihat seperti pada gambar berikut.



Gambar 2. 6 Menambahkan ikon Data Table

Lalu, membuat garis hubung dari ikon File ke ikon Data Tabel dengan cara mengklik lengkungan kanan pada ikon File, lalu arahkan ke lengkungan kiri pada ikon Data Table seperti pada gambar berikut.



Gambar 2. 7 Menyambungkan ikon File dengan ikon Data Table

Lalu, untuk melihat data yang sudah di input dapat dilakukan klik ganda pada ikon Data Table sehingga data dapat terlihat dalam bentuk tabel seperti pada gambar berikut.

Gambar 2. 8 Tampilan data dalam bentuk tabel

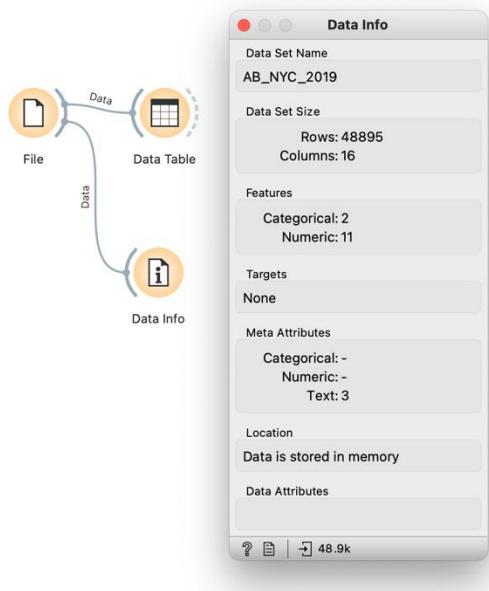
Dari tabel data tersebut, dapat diketahui bahwa terdapat 16 atribut yang dideteksi oleh Orange. Daftar atribut dari dataset AirBnB dapat dilihat pada tabel berikut.

Tabel 2. 1 Atribut *dataset* AirBnB

Atribut	Deskripsi
---------	-----------

name	Nama penginapan
host_name	Nama pemilik
neighbourhood	Lokasi penginapan
id	Kode unik penginapan
host_id	Kode unik pemilik
neighbourhood_group	Daerah lokasi penginapan
latitude	Koordinat garis lintang
longitude	Koordinat garis bujur
room_type	Tipe ruangan penginapan
price	Harga sewa penginapan
minimum_nights	Jumlah hari minimum penyewaan penginapan
number_of_review	Jumlah ulasan yang didapatkan penginapan
last_review	Tanggal terakhir penginapan diulas
reviews_per_month	Jumlah ulasan yang didapatkan penginapan dalam sebulan
calculated_host_listings_count	Jumlah penginapan yang dimiliki oleh satu pemilik dalam satu daerah
availability_365	Jumlah hari yang tersedia di penginapan dalam satu tahun

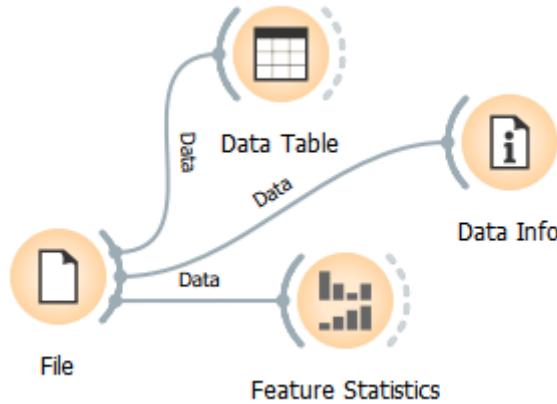
Info lebih lanjut mengenai dataset AirBnB dapat dilihat dengan fitur Data Info pada Orange seperti pada Gambar berikut.



Gambar 2. 9 Data Info

Pada Gambar di atas, ditampilkan bahwa Orange mendeteksi 16 atribut dengan 13 *features* yang terdiri dari 2 data *categorical* dan 11 data *numeric*. Sementara itu, 3 atribut sisanya adalah *meta-attributes* yang merupakan data *text*, yakni atribut *name*, *host\_name*, dan *neighbourhood*. Pada Gambar 2.55 juga ditampilkan bahwa dataset AirBnB ini memiliki 48,895 baris dan 13

*features* dengan 3,2% *missing data*, sehingga perlu dilakukan *data preprocessing* untuk kemudian dapat dilakukan analisis dengan data yang sudah bersih. Untuk melihat *missing data*, dapat menggunakan fitur *Feature Statistics* yang dihubungkan dengan ikon *File* seperti pada Gambar berikut.



Gambar 2. 10 Hubungkan *feature statistics*

Setelah dihubungkan, klik ganda pada ikon *Feature Statistics* dan ditampilkan halaman seperti pada Gambar berikut.



Gambar 2. 11 Info statistik data

### 2.2.3 Data Preprocessing

Dari Gambar di atas, *missing data* yang ditemukan pada dataset yaitu pada atribut *last\_review* dan *reviews\_per\_month* dengan *missing data* sebesar 20% untuk masing-masing atribut. *Missing data* pada kedua atribut tersebut dapat dilihat pada Gambar berikut.

Info  
48895 instances  
13 features (3.2 % missing data)  
No target variable.  
3 meta attributes (0.0 % missing data)

Variables  
 Show variable labels (if present)  
 Visualize numeric values  
 Color by instance classes

Selection  
 Select full rows

Restore Original Order

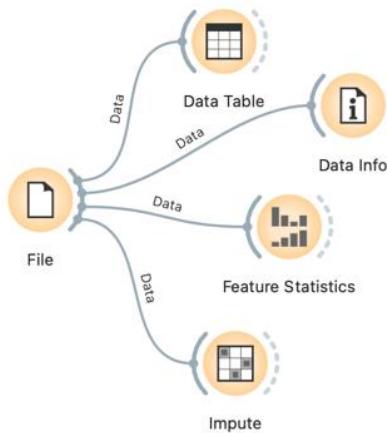
Send Automatically

Data Table

	price	minimum_nights	umber_of_review	last_review	views_per_month	ted_host_listings	availability_365
33138	1200	60	0	?	?	1	365
29171	70	15	0	?	?	1	0
47283	63	31	0	?	?	1	250
14020	150	1	0	?	?	1	0
27868	55	15	0	?	?	1	0
39743	1350	1	0	?	?	1	365
47941	52	1	0	?	?	3	33
47943	60	20	0	?	?	3	180
47942	54	1	0	?	?	3	6
40012	60	2	0	?	?	1	0
48489	30	7	0	?	?	1	190
27979	40	5	0	?	?	2	0
27064	90	5	0	?	?	2	0
47061	149	2	0	?	?	2	53
47423	149	2	0	?	?	2	23
44371	139	20	0	?	?	2	18
47833	250	1	0	?	?	2	17
47827	63	10	0	?	?	2	83
9786	150	1	0	?	?	1	0
47940	30	7	0	?	?	1	83

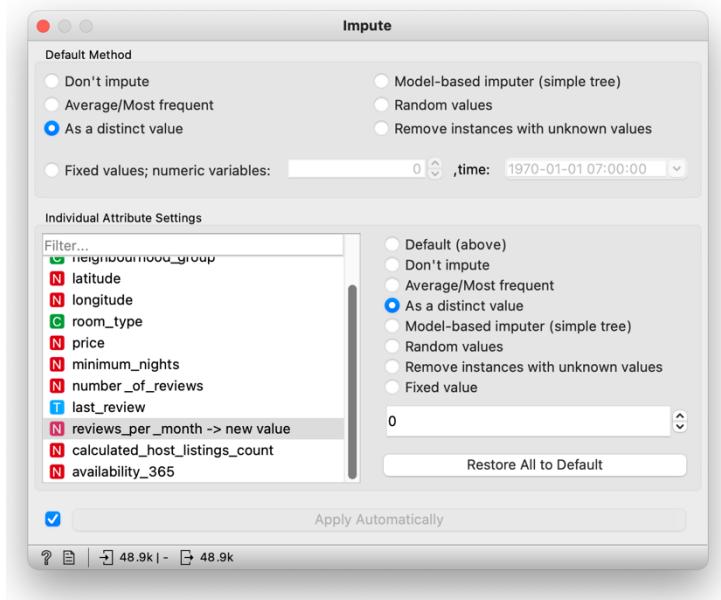
Gambar 2. 12 *missing data* pada kolom last\_review dan reviews\_per\_month

Untuk mengisi data yang kosong, dapat dilakukan *impute* dengan cara menghubungkan fitur *impute* ke *File* seperti pada gambar berikut.



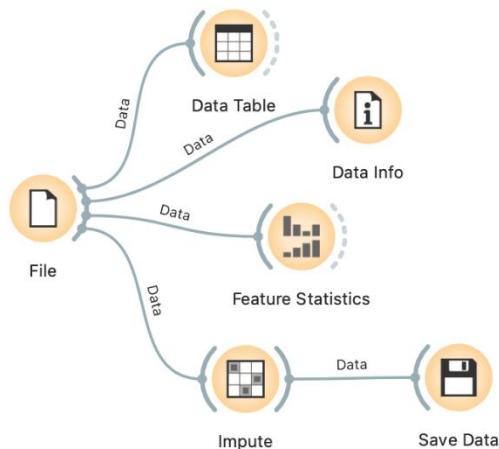
Gambar 2. 13 *Impute data*

Setelah dihubungkan, klik ganda ikon *Impute* dan kemudian ditampilkan halaman *Impute* seperti pada Gambar berikut. Untuk memberikan nilai pada data yang tidak lengkap pada atribut reviews\_per\_month, dapat memilih “*Impute as a distinct value*” dan diisi dengan nilai 0.

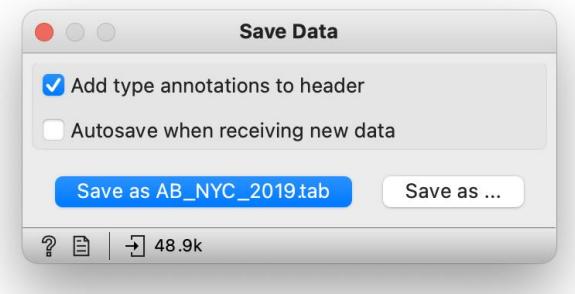


Gambar 2. 14 *Impute* data dengan nilai 0

Setelah data yang tidak memiliki nilai diisi, *dataset* yang baru harus disimpan dengan memilih fitur *Save Data* dan dihubungkan ke fitur *Impute* seperti pada gambar berikut.

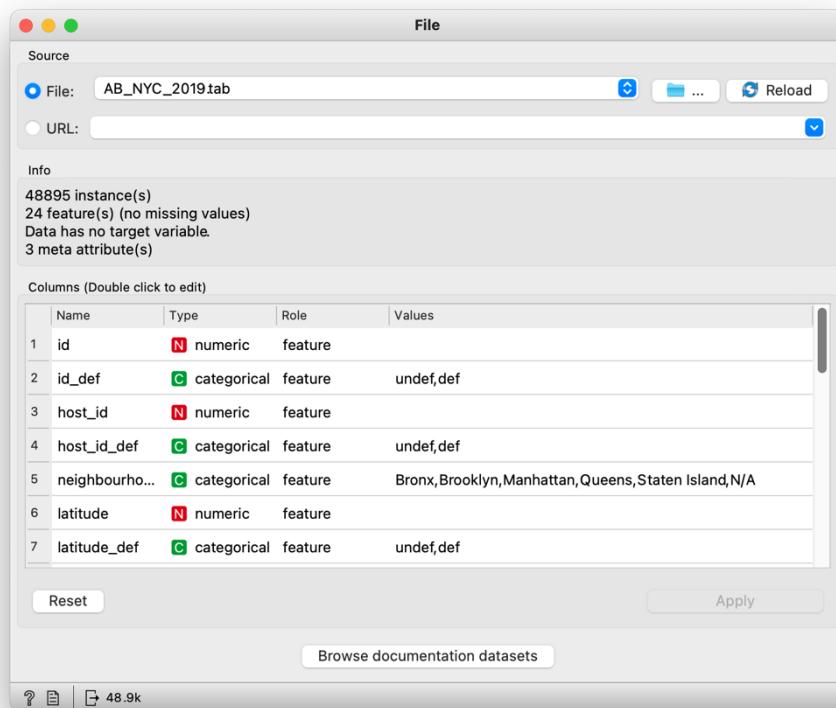


Gambar 2. 15 Hubungkan *save data*

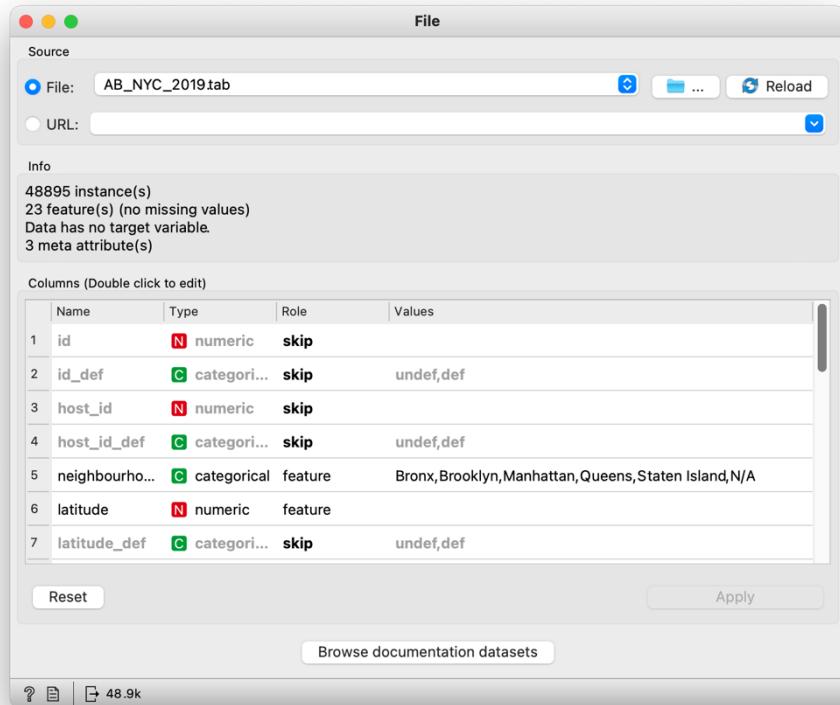


Gambar 2. 16 *Save data*

Fitur *Save Data* dapat diklik ganda dan ditampilkan halaman untuk menyimpan *dataset* yang baru dengan nama *file AB\_NYC\_2019.tab* pada gambar berikut yang menunjukkan perubahan *role* pada atribut yang tidak diperlukan.



Gambar 2. 17 File data



Gambar 2. 18 File data

Dataset yang baru sudah bersih dan dapat digunakan untuk analisis. Dari *Data Table* pada Gambar di atas dapat diketahui bahwa sudah tidak terdapat *missing data*. Pada Gambar berikut ditunjukkan bahwa atribut *reviews\_per\_month* sudah diisi dengan nilai 0.

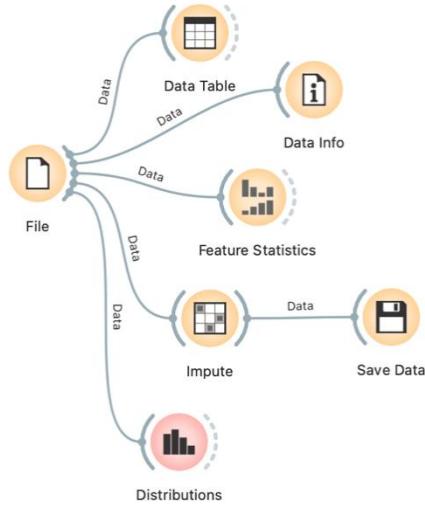
Data Table								
Info	room_type	price	minimum_nights	umber_of_review	last_review	reviews_per_mor~	ted_host_listings	availability_365
48895 instances	rivate room	90.0	7.0	0.0	2018-10-04	0.00	1.0	23.0
13 features	hared room	55.0	1.0	0.0	2018-10-04	0.00	6.0	2.0
No target variable	ntire home/...	115.0	10.0	0.0	2018-10-04	0.00	1.0	27.0
3 meta attributes (0.0 % missing data)	rivate room	40.0	4.0	0.0	2018-10-04	0.00	2.0	36.0
Variables	private room	70.0	2.0	0.0	2018-10-04	0.00	2.0	9.0
<input checked="" type="checkbox"/> Show variable labels (if present)	rivate room	65.0	1.0	0.0	2018-10-04	0.00	2.0	163.0
<input checked="" type="checkbox"/> Visualize numeric values	rivate room	125.0	4.0	0.0	2018-10-04	0.00	1.0	31.0
<input checked="" type="checkbox"/> Color by instance classes	ntire home/...	170.0	1.0	0.0	2018-10-04	0.00	3.0	365.0
Selection	ntire home/...	200.0	6.0	0.0	2018-10-04	0.00	1.0	176.0
<input checked="" type="checkbox"/> Select full rows	rivate room	75.0	2.0	0.0	2018-10-04	0.00	2.0	353.0
	rivate room	190.0	7.0	0.0	2018-10-04	0.00	1.0	341.0
	rivate room	75.0	2.0	0.0	2018-10-04	0.00	2.0	364.0
	rivate room	40.0	20.0	0.0	2018-10-04	0.00	1.0	31.0
	rivate room	54.0	6.0	0.0	2018-10-04	0.00	1.0	15.0
	ntire home/...	120.0	1.0	0.0	2018-10-04	0.00	3.0	365.0
	ntire home/...	120.0	20.0	0.0	2018-10-04	0.00	1.0	22.0
	rivate room	45.0	1.0	0.0	2018-10-04	0.00	5.0	172.0
	rivate room	42.0	7.0	0.0	2018-10-04	0.00	1.0	16.0
	rivate room	60.0	1.0	0.0	2018-10-04	0.00	1.0	26.0
	.....	140.0	1.0	0.0	2018-10-04	0.00	1.0	180.0

Gambar 2. 19 Data table bersih

## 2.2.4 Data Exploratory & Distribution

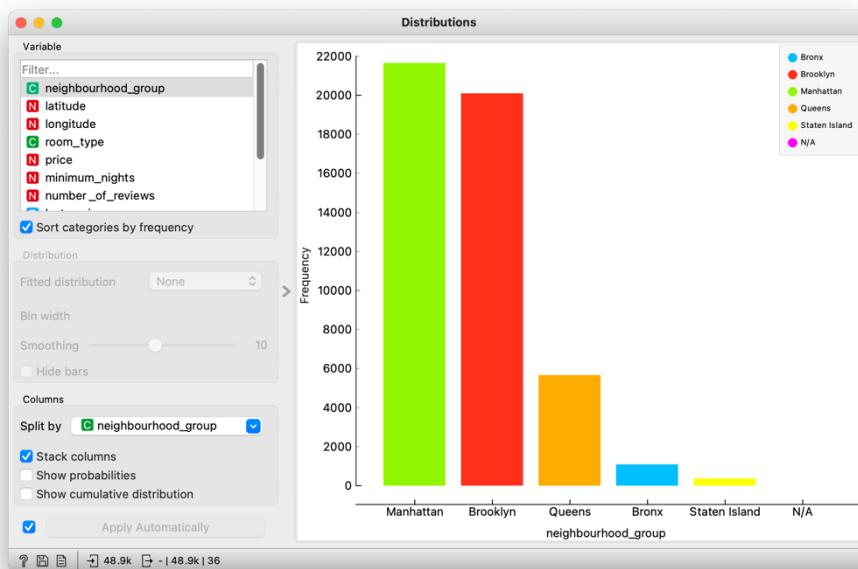
Dataset yang sudah bersih setelah dilakukan *pre-processing* dapat digunakan untuk melakukan analisis. Untuk dapat melakukan analisis, dilakukan eksplorasi data dengan data *categorical*

dan *numerical*. Untuk data *categorical*, dapat menggunakan fitur *Distribution* yang dihubungkan ke ikon File untuk melihat persebaran data seperti pada Gambar berikut.



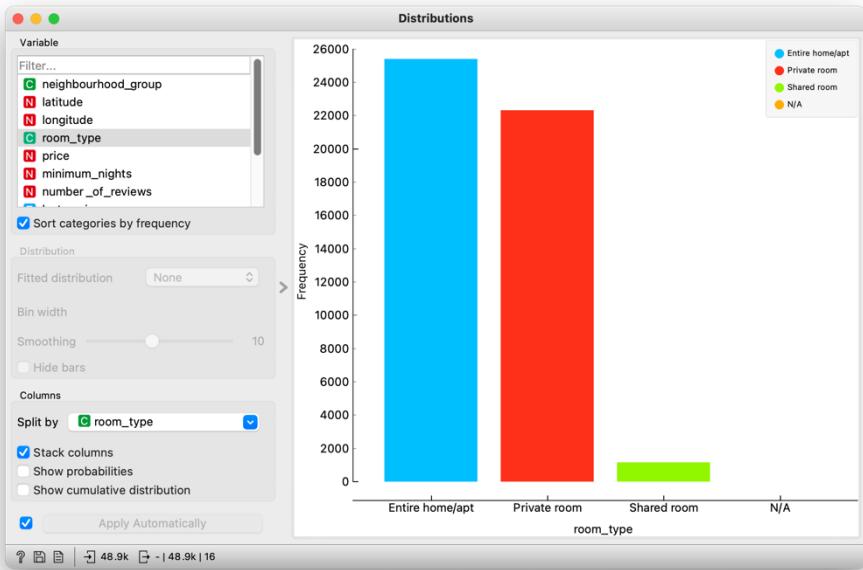
Gambar 2. 20 Hubungkan *distributions*

Dengan fitur *Distribution*, dapat dilihat persebaran data untuk masing-masing nilai pada setiap atribut. Pada Gambar 2.x menunjukkan persebaran data *neighbourhood\_group*.



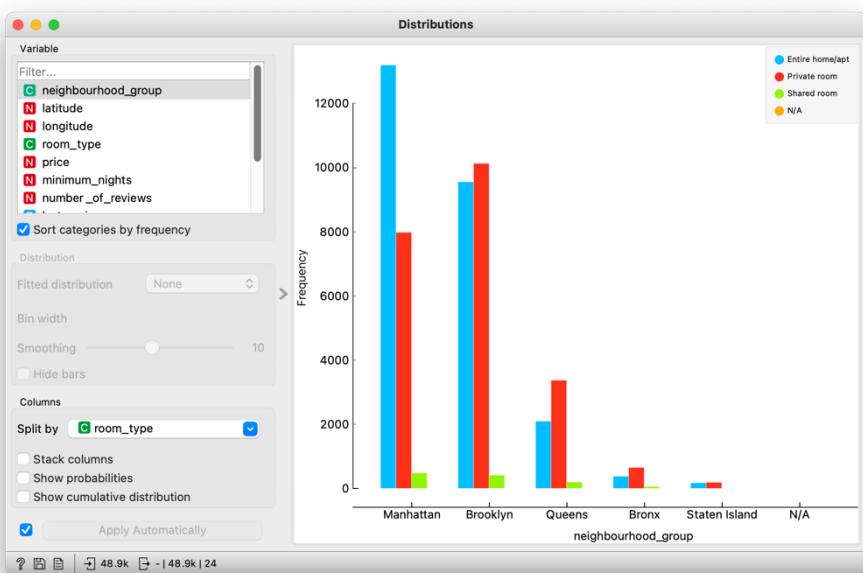
Gambar 2. 21 data *Neighbourhood group*

Dari keseluruhan penginapan, ditunjukkan bahwa sebagian besar penginapan terdapat pada daerah Manhattan sebesar 44% dan Brooklyn sebesar 41%. Dari hasil ini dapat diketahui bahwa kedua daerah ini merupakan daerah strategis, sehingga terdapat banyak penginapan. Selanjutnya, pada Gambar berikut menunjukkan persebaran data *room\_type*.



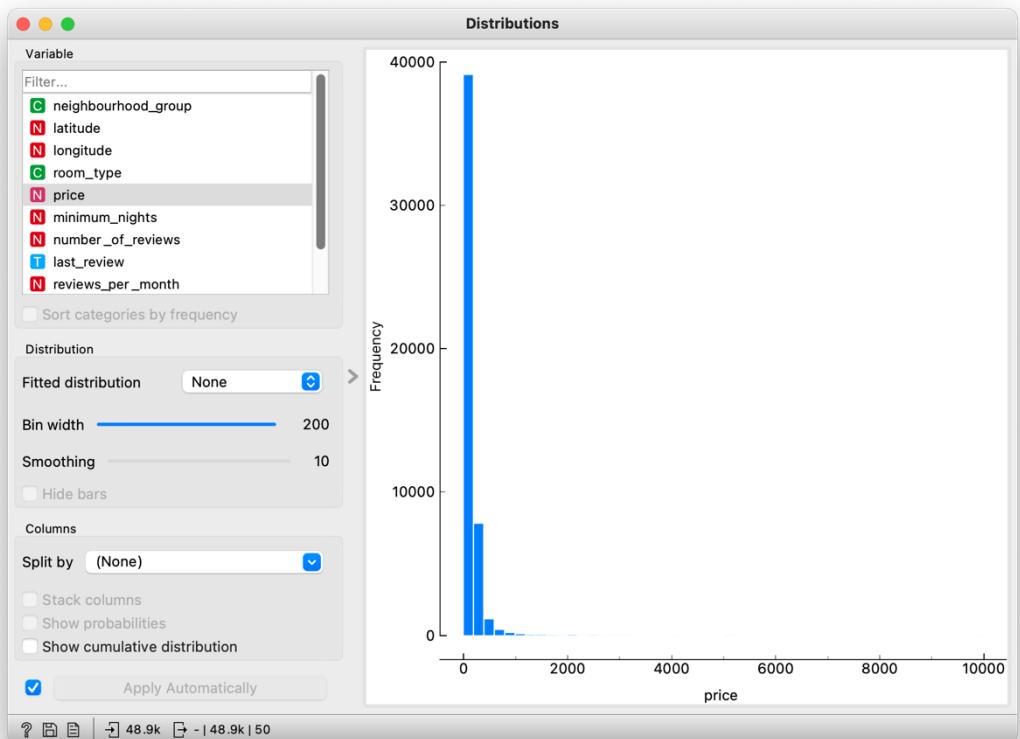
Gambar 2. 22 data *room type*

Dari keseluruhan penginapan, ditunjukkan bahwa sebagian besar penginapan merupakan tipe satu rumah atau apartemen sebesar 52% dan tipe *private room* sebesar 45%. Selanjutnya, pada Gambar berikut menunjukkan persebaran data *room\_type* dan *neighbourhood\_group*.



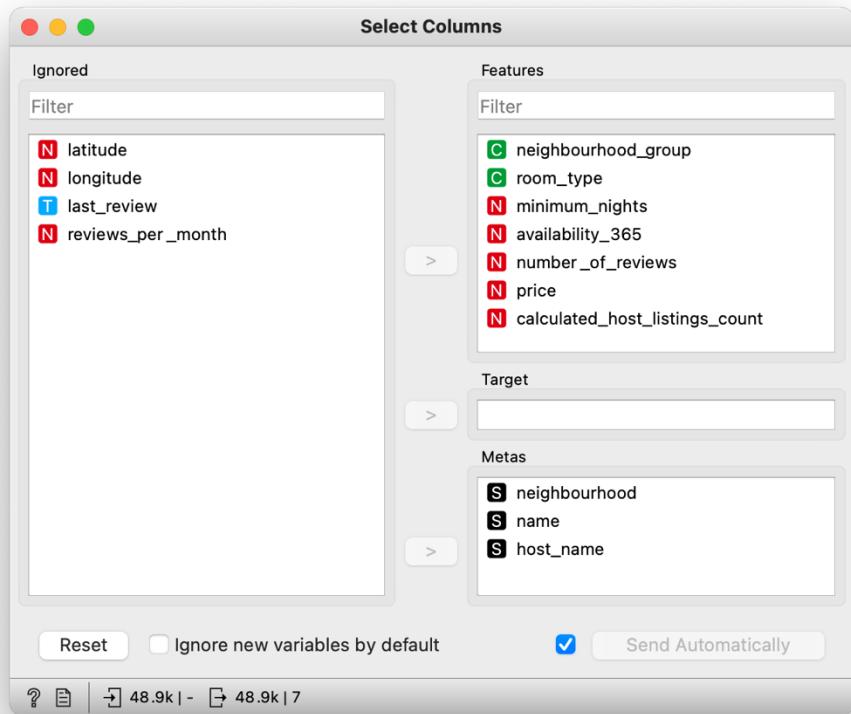
Gambar 2. 23 persebaran data *room\_type* dan *neighbourhood\_group*

Dari keseluruhan penginapan, ditunjukkan bahwa Manhattan memiliki penginapan tipe rumah atau apartemen yang secara signifikan lebih tinggi dibandingkan tipe lainnya. Di Brooklyn, penginapan tipe rumah atau apartemen dan *private room* memiliki jumlah yang tidak berbeda jauh, sedangkan di daerah lain didominasi oleh penginapan tipe *private room*. Selanjutnya, Gambar berikut menunjukkan persebaran data *price*.

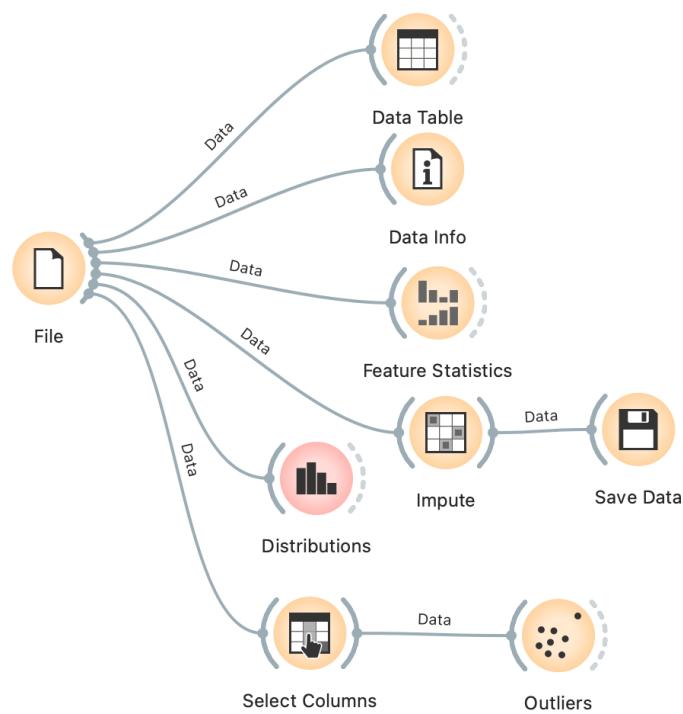


Gambar 2. 24 Data price

Dari data *price*, ditunjukkan bahwa persebaran data berkumpul pada satu area. Hal ini dikarenakan dengan jangkauan nilai data yang ada, terdapat beberapa nilai yang bernilai jauh dengan data lainnya dengan jumlah yang tidak terlalu banyak. Hal ini biasa disebut sebagai *outliers*, sehingga perlu dilakukan *outlier detection* dengan fitur *Outliers* pada Orange yang dihubungkan dengan fitur *Select Columns* untuk menentukan atribut yang akan digunakan seperti pada gambar berikut.

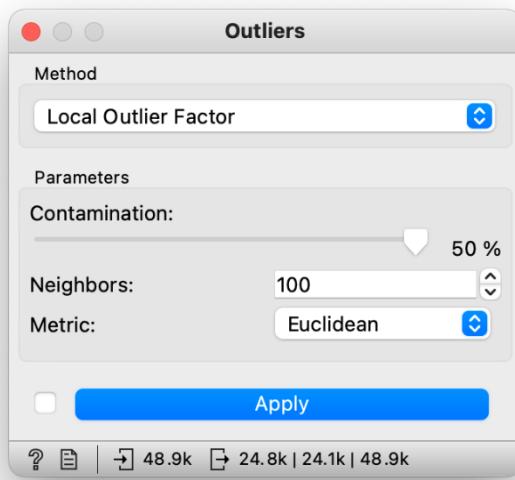


Gambar 2. 25 Fitur *Select columns*



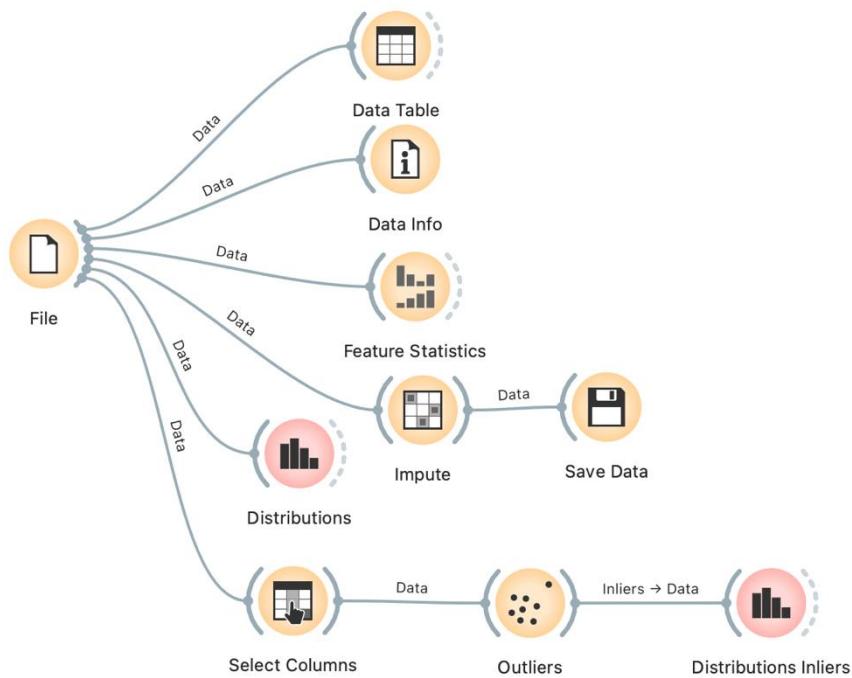
Gambar 2. 26 Hubungkan fitur *outliers*

Setelah dihubungkan, klik ganda ikon *Outliers* dan kemudian ditampilkan halaman seperti pada Gambar berikut.



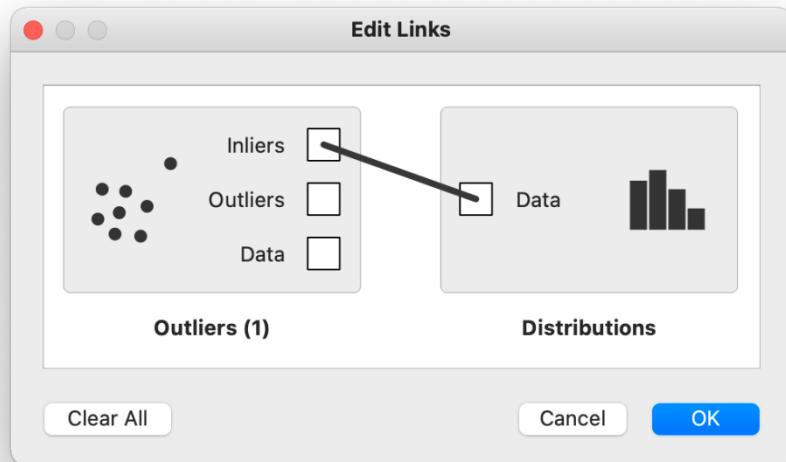
Gambar 2. 27 Pengaturan *outliers*

Untuk pengaturan *outlier detection* yang dipilih adalah dengan *contamination* sebesar 50% dan *neighbors* sebesar 100. Setelah selesai dilakukan pengaturan, fitur *Outliers* dihubungkan dengan fitur *Distribution* untuk melihat persebaran data seperti pada gambar.



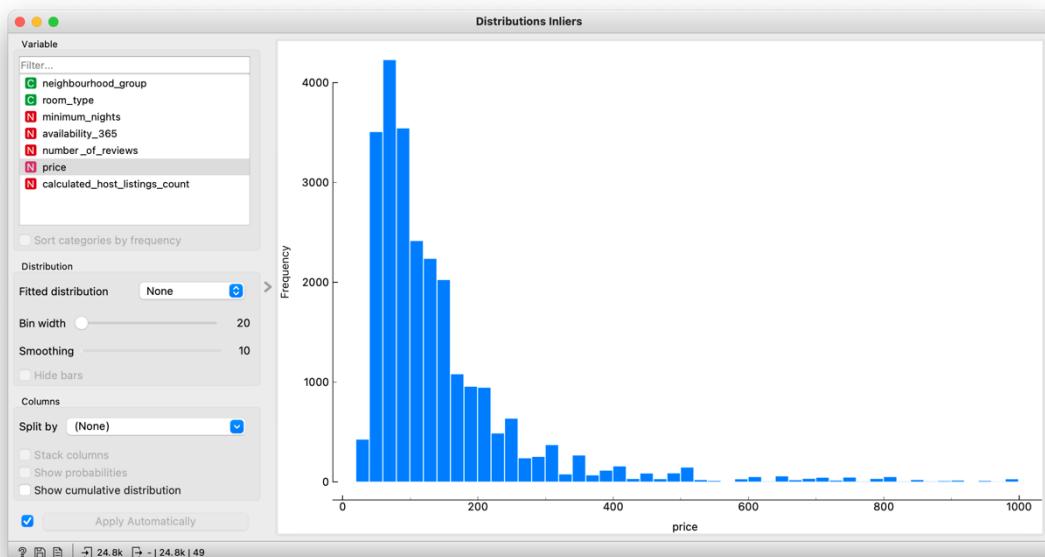
Gambar 2. 28 Hubungkan fitur *Distributions*

Setelah dihubungkan, ditampilkan halaman untuk memilih data yang akan ditampilkan pada fitur *Distributions*. Untuk hal ini, data yang dipilih untuk digunakan adalah data *inliers* seperti pada Gambar berikut.



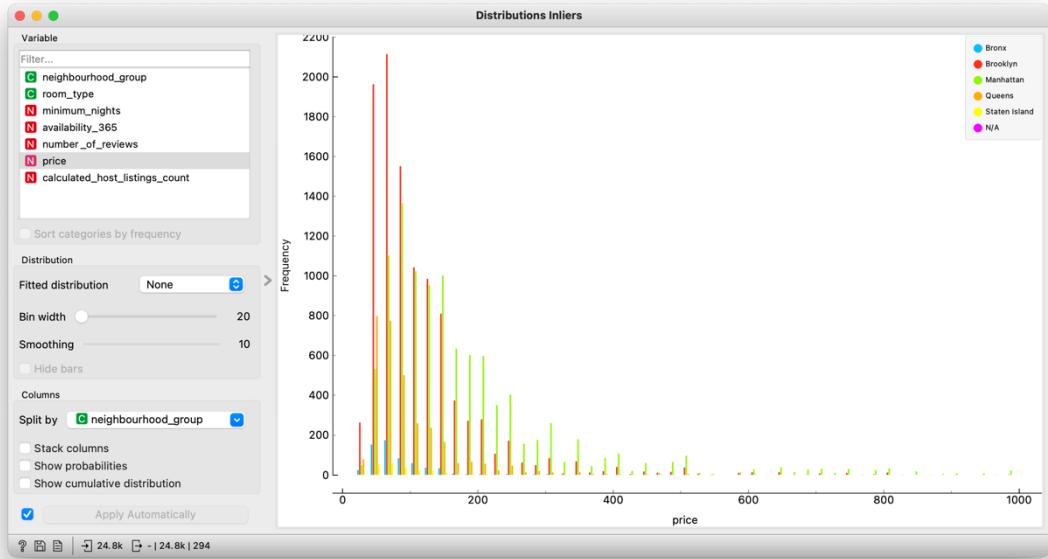
Gambar 2. 29 Pemilihan data untuk *distributions*

Persebaran data setelah dihilangkan *outliers* terlihat pada Gambar berikut.



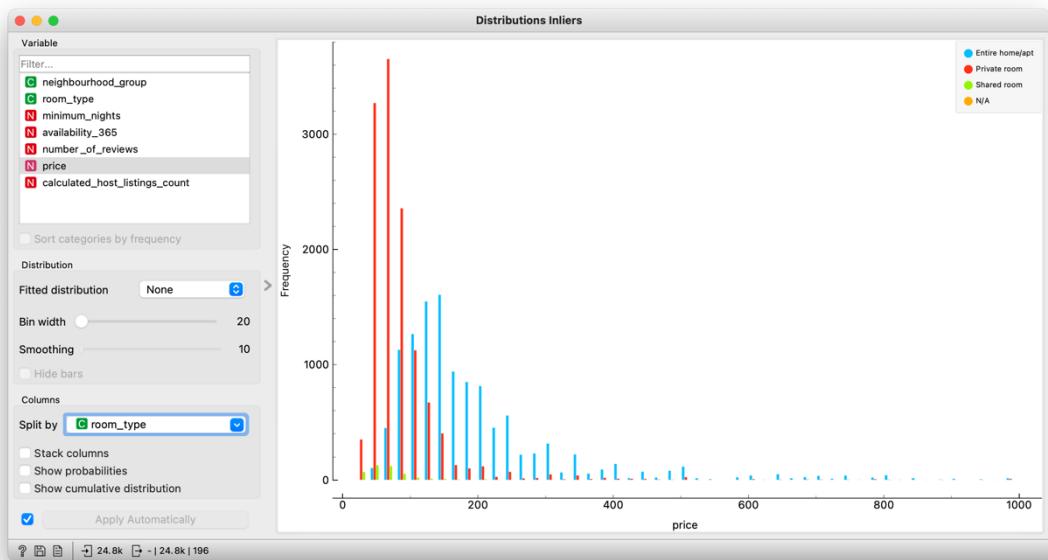
Gambar 2. 30 Persebaran *data price inliers*

Data *price* dan *neighbourhood\_group* ditampilkan pada Gambar berikut.



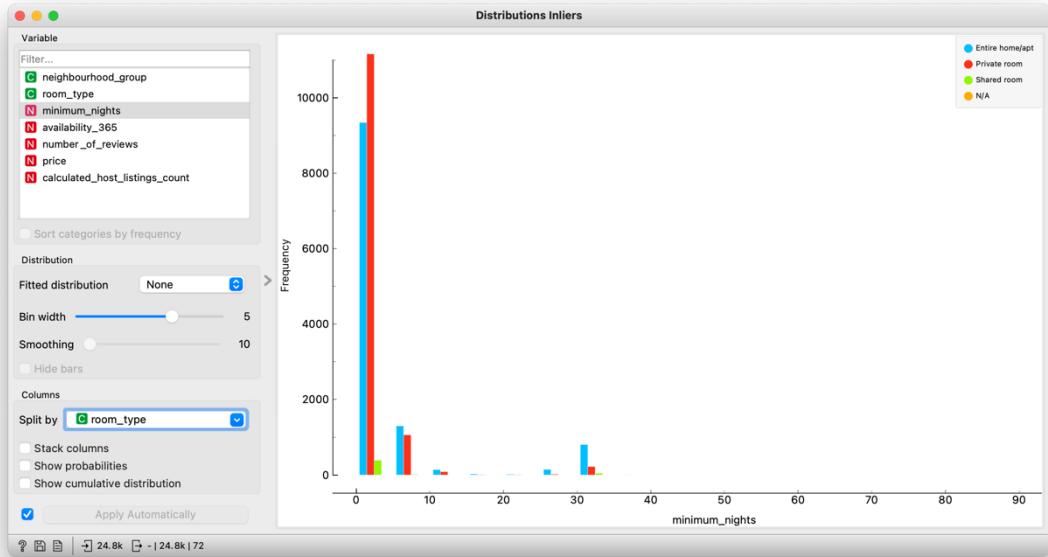
Gambar 2. 31 Persebaran *data price* dan *neighbourhood group*

Hasil pada Gambar di atas menunjukkan bahwa penginapan di Brooklyn memiliki harga yang relatif lebih tinggi dari daerah lainnya, kemudian diikuti Manhattan. Selanjutnya, gambar berikut menunjukkan data *price* dan *room\_type*.



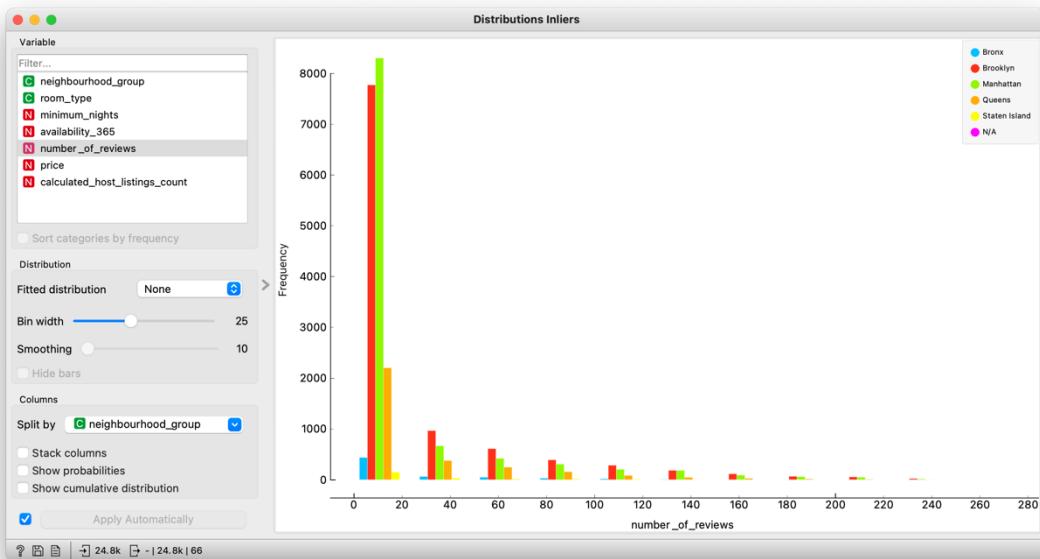
Gambar 2. 32 Persebaran *data price* dan *room type*

Gambar di atas menunjukkan bahwa penginapan dengan tipe *private room* cenderung memiliki harga yang relatif lebih tinggi jika dibandingkan dengan tipe lainnya, seperti tipe rumah dan apartemen. Selanjutnya, gambar berikut menunjukkan data *minimum\_nights* dan *room\_type*.



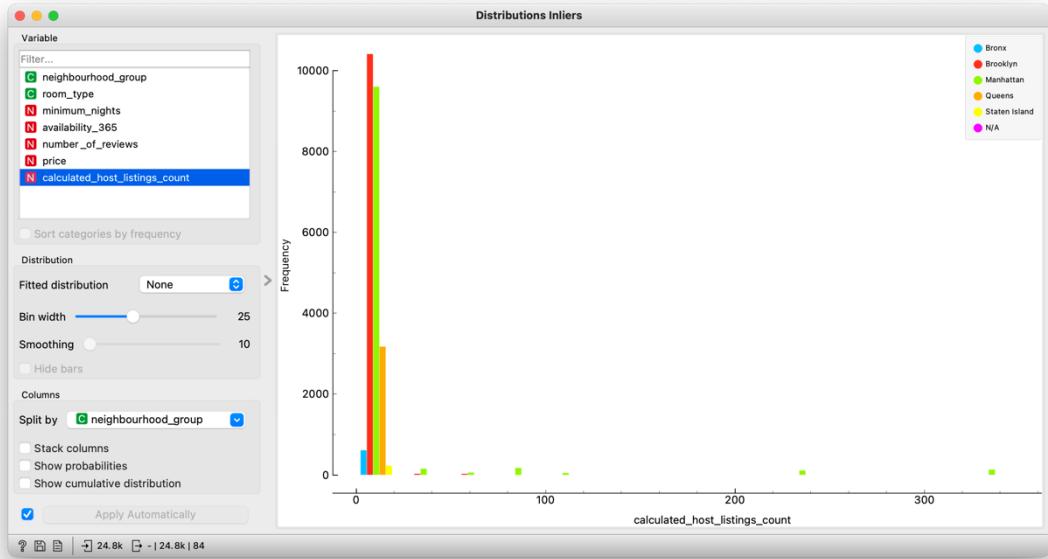
Gambar 2. 33 Persebaran *data minimum nights* dan *room type*

Gambar di atas menunjukkan bahwa sebagian besar tipe penginapan memiliki jumlah hari sewa minimum yang mirip, yaitu kurang dari 5 hari. Namun, untuk tipe rumah atau apartemen juga cenderung memiliki jumlah hari sewa minimum sebesar 30 hari atau 1 bulan. Selanjutnya, gambar berikut menunjukkan data *number\_of\_reviews* dan *neighbourhood\_group*.



Gambar 2. 34 Persebaran *data number of reviews* dan *neighbourhood group*

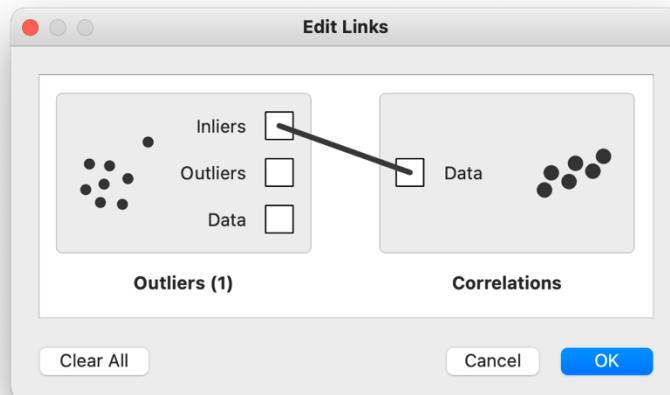
Gambar di atas menunjukkan bahwa penginapan di Brooklyn dan Manhattan memiliki jumlah ulasan yang lebih tinggi dibandingkan dengan daerah lainnya. Hal ini menunjukkan bahwa Brooklyn dan Manhattan seringkali menjadi pilihan daerah penginapan. Selanjutnya, gambar berikut menunjukkan data *calculated\_host\_listings* dan *neighbourhood\_group*.



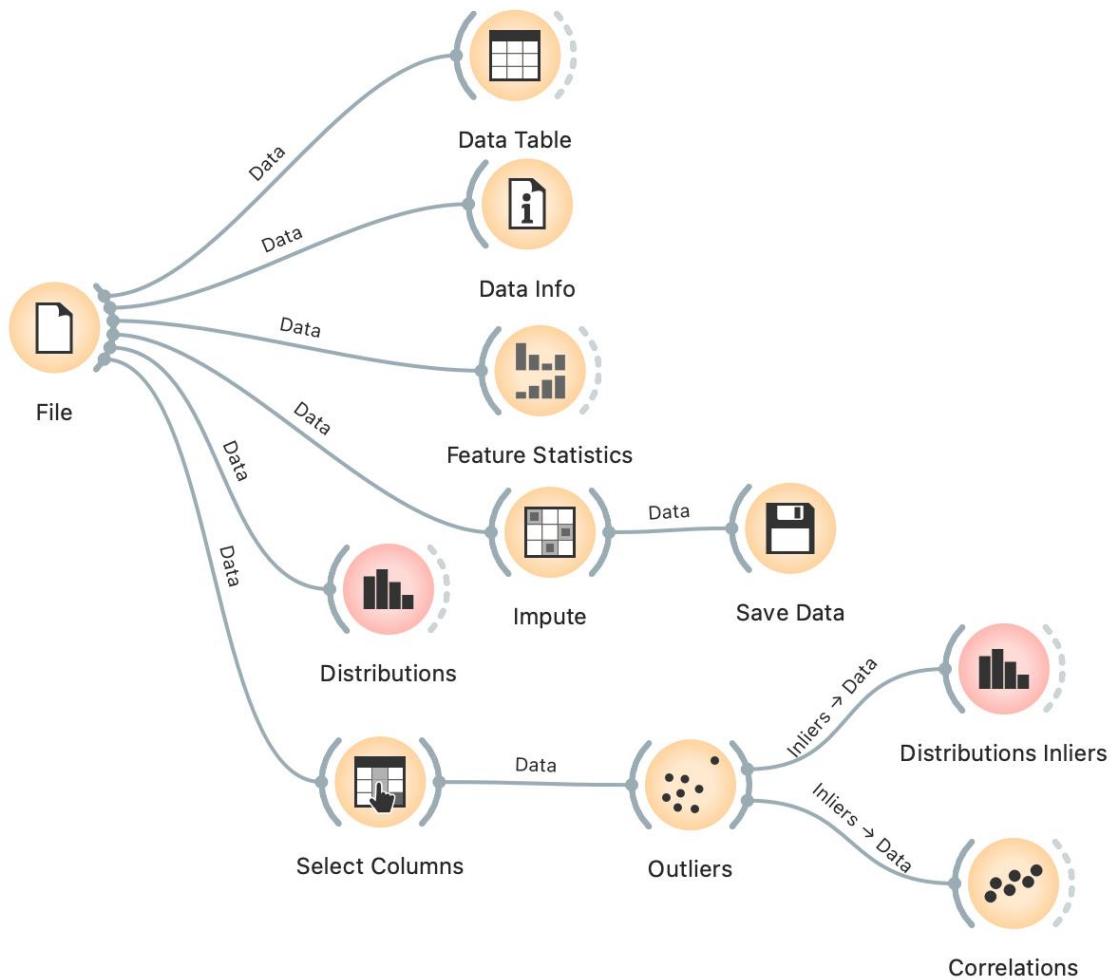
Gambar 2. 35 Persebaran *data host listings* dan *neighbourhood group*

Gambar di atas menunjukkan bahwa pemilik penginapan yang memiliki jumlah penginapan yang banyak berada pada daerah Brooklyn dan Manhattan. Hal ini mengindikasikan bahwa *host listing* dengan nilai yang tinggi adalah *host* yang sudah berpengalaman. Maka dari itu, penginapan di Brooklyn dan Manhattan diketahui memiliki penginapan yang lebih baik.

Setelah diketahui persebaran data, dilihat juga hubungan data dengan fitur *Correlations* seperti pada gambar berikut.

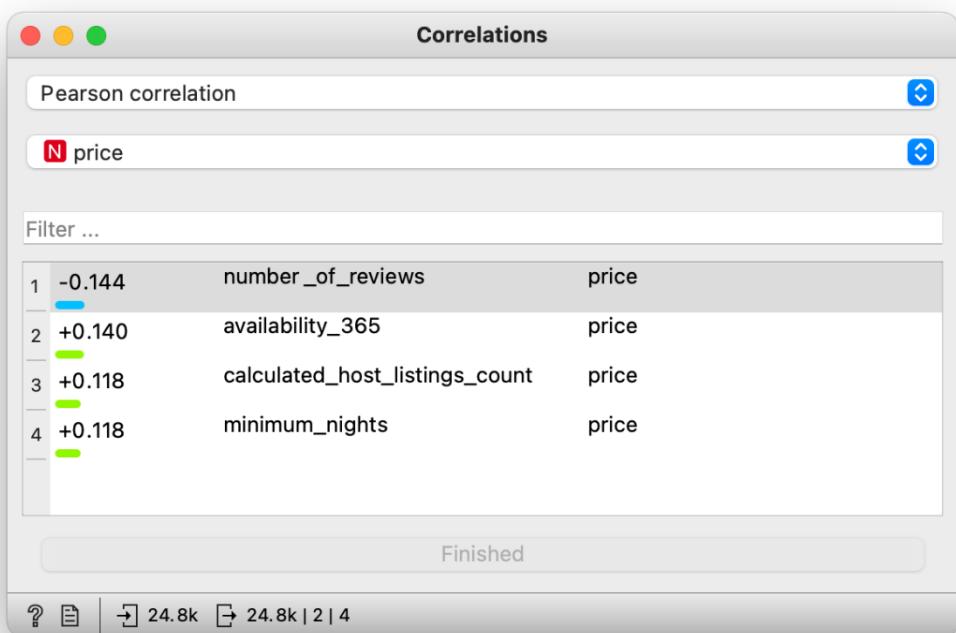


Gambar 2. 36 Pengaturan *correlations*



Gambar 2. 37 Hubungkan *correlations*

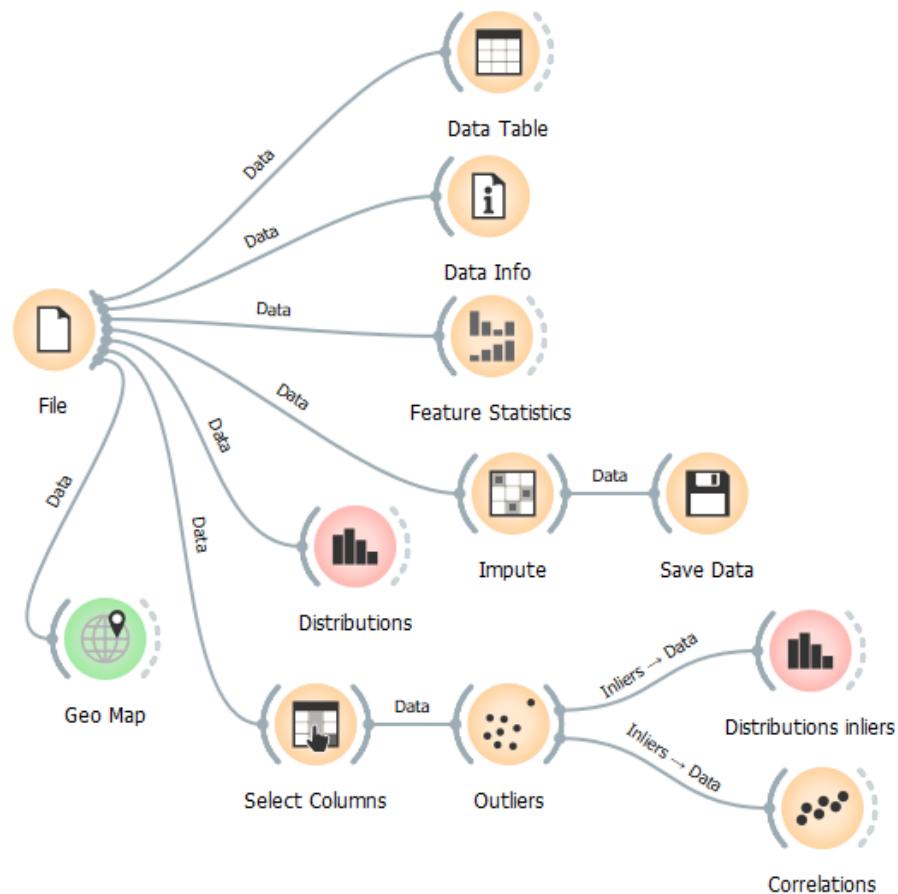
Hasil untuk hubungan data dengan atribut *price* adalah sebagai berikut.



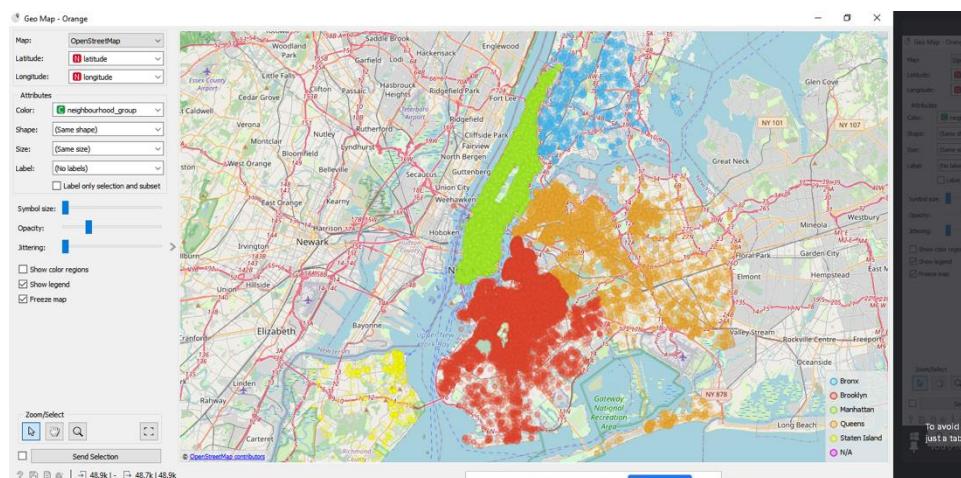
Gambar 2. 38 Hubungan Korelasi data

Gambar tersebut menunjukkan bahwa harga memiliki hubungan yang kuat dengan jumlah ketersediaan hari sewa yang dapat diindikasikan bahwa semakin tinggi jumlah ketersediaan penginapan, maka bisa semakin tinggi juga harga yang diberikan. Hubungan yang kuat juga sama antara harga dengan jumlah kepemilikan dari *hosting*, yang diindikasikan bahwa semakin tinggi jumlah penginapan yang dimiliki oleh *hosting*, maka semakin tinggi juga harga yang diberikan, melihat bahwa *hosting* tersebut sudah berpengalaman dalam memberikan layanan penginapan.

Selanjutnya, data persebaran penginapan AirBnB di New York berdasarkan wilayah dapat dilihat menggunakan fitur *Geo Map* yang dihubungkan dengan ikon *File*. Persebaran data *map* ini menggunakan atribut *latitude* dan *longitude*. Hasil yang didapatkan adalah sebagai berikut.



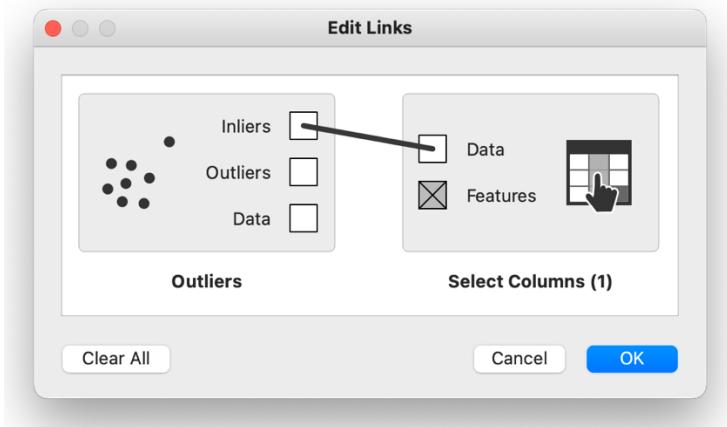
Gambar 2. 39 Hubungkan fitur *geo map*



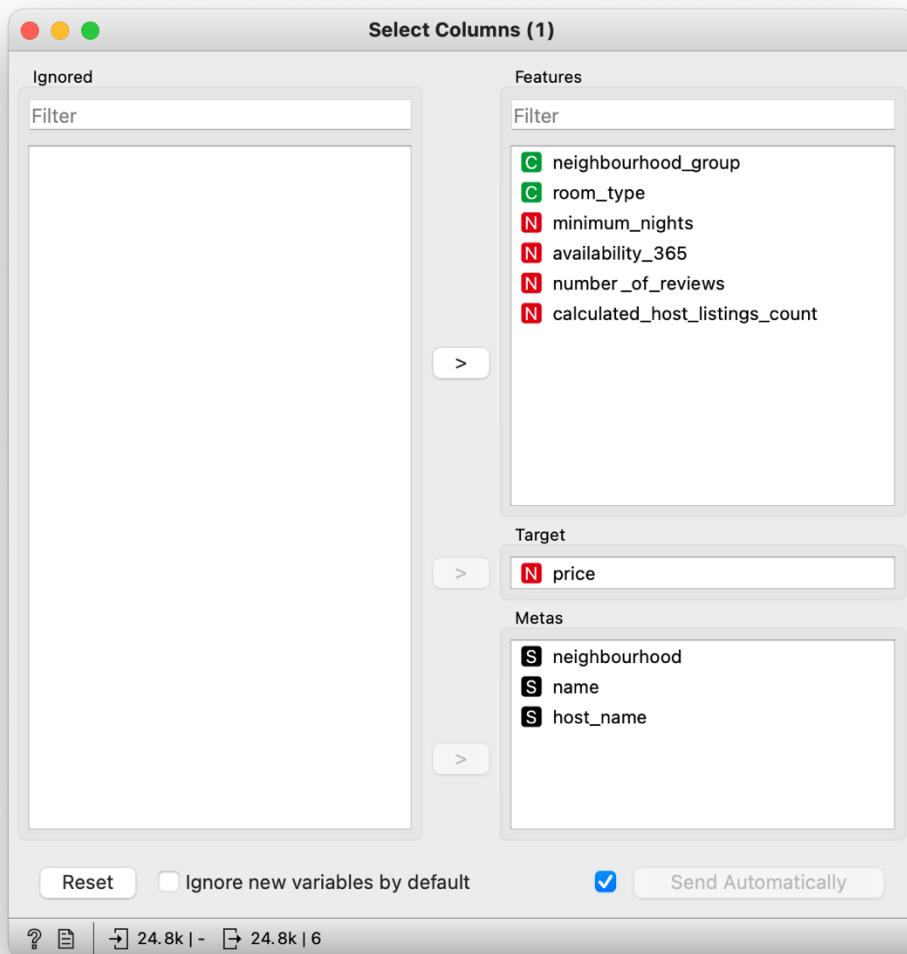
Gambar 2. 40 Persebaran data berdasarkan daerah

## 2.2.5 Data Modelling

Pemodelan untuk data AB\_NYC\_2019 ini dapat dilakukan dengan memilih atribut yang digunakan terlebih dahulu dan menjadikan *price* sebagai *target*. Atribut yang dipilih adalah sebagai berikut.

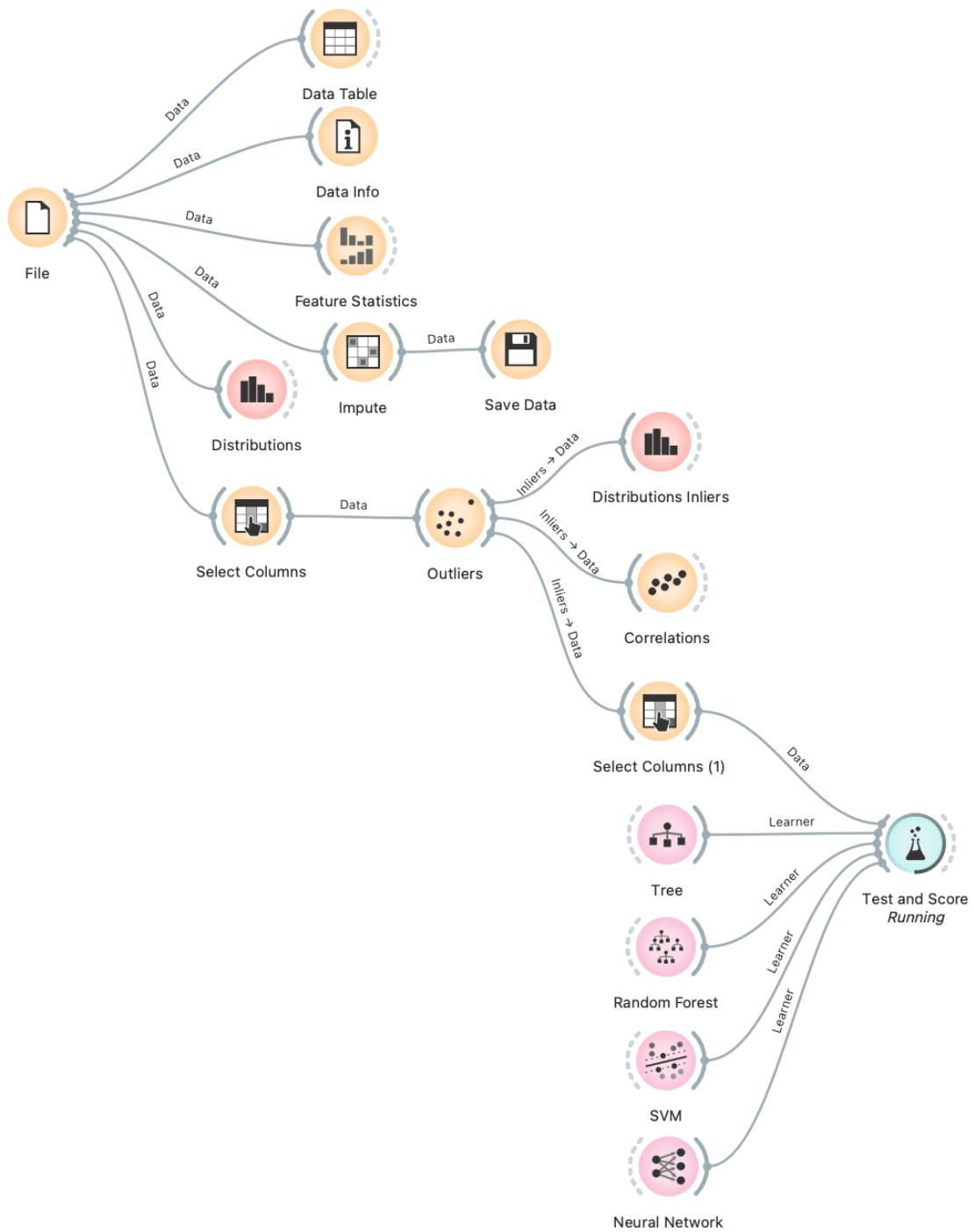


Gambar 2. 41 Pengaturan data untuk *select columns*



Gambar 2. 42 *Select columns*

Untuk melakukan pemodelan data, fitur masing-masing model klasifikasi dapat dimasukan ke dalam lembar kerja Orange seperti pada Gambar berikut.



Gambar 2. 43 Pemodelan data

Pada Gambar di atas menunjukkan bahwa model klasifikasi yang digunakan untuk pemodelan data ini adalah sebagai berikut.

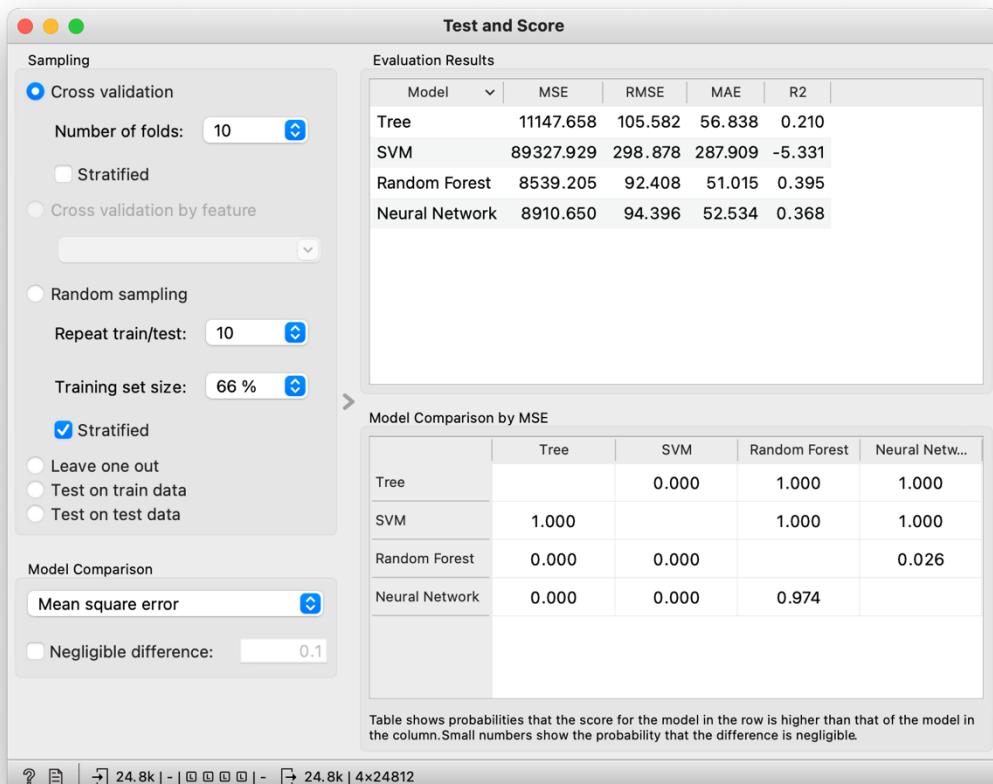
- *Support Vector Machine (SVM)*
- *Neural Network*
- *Random Forest*

- *Decision Tree*

Setelah itu, dilakukan evaluasi dengan menggunakan fitur *Test and Score* pada Orange.

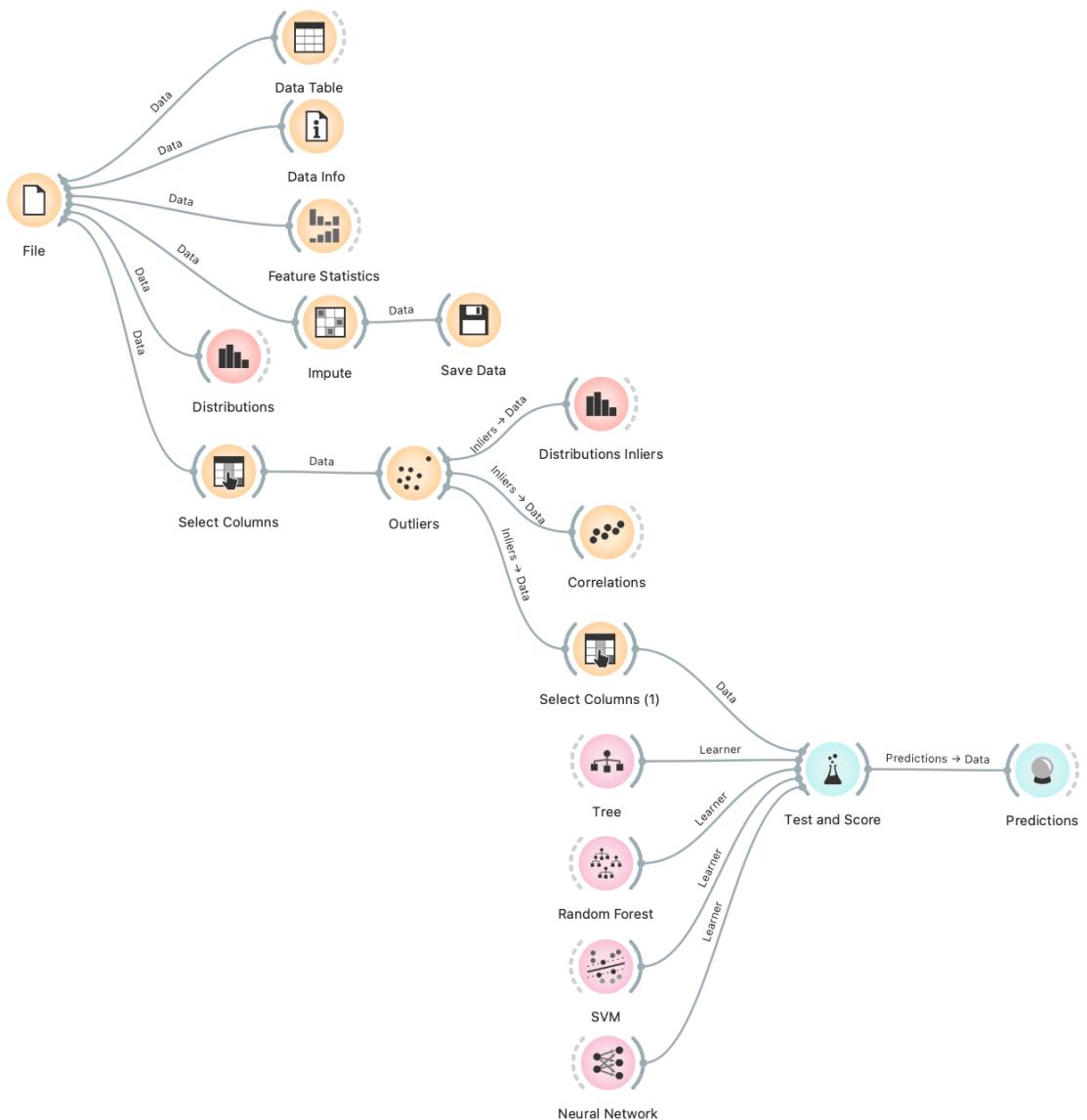
## 2.2.6 Results and Analysis

Dari pemodelan data yang sudah dilakukan, kemudian dilakukan evaluasi dengan fitur *Test and Score* dan didapatkan hasil seperti pada Gambar berikut.



Gambar 2. 44 Hasil evaluasi

Pada Gambar di atas dapat dilihat pengaturan fitur *Test and Score* dengan menggunakan teknik *sampling k-fold cross validation* dengan nilai k sebesar 10. Hasil yang didapatkan menunjukkan bahwa *Random Forest* adalah algoritme yang mendapatkan hasil kinerja tertinggi dibandingkan dengan algoritme lainnya, yakni dengan nilai MSE sebesar 8.539. Dari hasil perhitungan kinerja model tersebut, fitur *Predictions* dihubungkan dengan *Test and Score* dan menampilkan hasil *predictions* yang didapatkan adalah sebagai berikut.



Gambar 2. 45 Hubungkan *Predictions*

Predictions

Show probabilities for

Restore Original Order

price	neighbourhood	name	host_name	Tree	Random Forest	SVM	Neural Network	Fold	neighborhood
110.0	Lower East S...	Great Locati...	Bianca	148.75	123.128	413.329	125.92	1	Manhi
49.0	Greenpoint	Room in Gre...	Andrea	47.5	67.9771	413.569	80.4386	1	Brookl
99.0	East Village	Large Room i...	Casey	73.5	100.161	413.473	90.6638	1	Manhi
85.0	Washington ...	HARLEM,NE...	Tye And Etie...	130.5	131.106	413.368	111.707	1	Manhi
60.0	East Village	Oh glorious ...	Edward	73.5	87.3848	413.599	89.5167	1	Manhi
125.0	Bedford-Stu...	The Stuydio ...	Shelly	147	104.905	413.584	117.698	1	Brookl
140.0	Flushing	Fully Furnish...	Fredah	78.75	60.6412	413.538	69.2867	1	Queer
156.0	West Village	Sun Filled Cl...	Karen	85	97.1167	413.41	82.4036	1	Manhi
90.0	Bedford-Stu...	SAFE AND B...	Marilyn	50	53.6938	413.562	79.4268	1	Brookl
45.0	Highbridge	Cozy bedroo...	Waundell	81.6667	62.9594	413.642	53.0979	1	Bronx
69.0	Williamsburg	Spacious Wil...	Gus	103.75	75.5071	413.513	73.826	1	Brookl
99.0	Bushwick	BROWNSTO...	Sheila	87.5	135.052	413.548	111.776	1	Brookl
75.0	Inwood	HOSTING Y...	Paula	54.5	87.9926	413.343	166.954	1	Manhi
129.0	East Village	CHARMING ...	Mary D	85	155.819	413.429	288.039	1	Manhi
190.0	Gowanus	LovelyMode...	Jana	149.333	135.85	413.524	133.234	1	Brookl
350.0	Clinton Hill	Lovely 3 bed...	Bliss	142.5	204.078	413.511	184.112	1	Brookl
185.0	Bedford-Stu...	BEAUTIFUL ...	Ruthven	142.5	143.639	413.511	142.976	1	Brookl
65.0	Harlem	Sun-Drench...	Jane	48.5	70.8017	413.456	68.9296	1	Manhi
80.0	Crown Heigh...	ROOM WITH...	Marilyn	89.5	127.133	413.629	126.456	1	Brookl
495.0	Park Slope	Architect's B...	Orna	103.75	285.033	413.534	202.315	1	Brookl
200.0	Upper West ...	LOCATION L...	Ligia	162	175.735	413.476	185.467	1	Manhi
140.0	SoHo	*SoHo: Clea...	Jade	100	76.5328	413.353	88.634	1	Manhi
500.0	Carroll Gard...	Amazing Bro...	Smadar	250	207.025	413.545	173.477	1	Brookl
75.0	Bedford-Stu...	Warehouse ...	Julia	73.5	83.1927	413.539	93.0436	1	Brookl
79.0	East Harlem	The Happy h...	Lena	110.2	112.295	413.471	171.914	1	Manhi
500.0	Greenpoint	Brand New ...	Sarah	148	138.286	413.502	164.235	1	Brookl

24.8k | - 24.8k | -

Gambar 2. 46 Hasil predictions