

پیاده سازی

قسمت اول :

جدول داده ها به صورت زیر می باشد.

id	sex	birth_year	country	region	infection_reason	infected_by	confirmed_date	state
1	female	1984	China	filtered at airport	visit to Wuhan		#####	released
2	male	1964	Korea	filtered at airport	visit to Wuhan		#####	released
3	male	1966	Korea	capital area	visit to Wuhan		#####	released
4	male	1964	Korea	capital area	visit to Wuhan		#####	released
5	male	1987	Korea	capital area	visit to Wuhan		#####	released
6	male	1964	Korea	capital area	contact with patient	3	#####	released
7	male	1991	Korea	capital area	visit to Wuhan		#####	released
8	female	1957	Korea	Jeollabuk-do	visit to Wuhan		#####	released
9	female	1992	Korea	capital area	contact with patient	5	2/1/2020	released
10	female	1966	Korea	capital area	contact with patient	6	2/2/2020	released
11	male	1995	Korea	capital area	contact with patient	6	2/3/2020	released
12	male	1971	China	capital area	contact with patient in Japan		2/4/2020	released
13	male	1992	Korea	filtered at airport	residence in Wuhan		2/5/2020	released
14	female	1980	China	capital area	contact with patient	12	2/6/2020	released
15	male	1977	Korea	capital area	contact with patient	4	2/7/2020	released
16	female	1977	Korea	Gwangju	visit to Thailand		2/8/2020	released
17	male	1982	Korea	capital area	contact with patient in Singapore		2/9/2020	released
18	female	1999	Korea	Gwangju	contact with patient	16	#####	released
19	male	1983	Korea	capital area	contact with patient in Singapore		#####	released
20	female	1978	Korea	capital area	contact with patient	15	#####	released
21	female	1960	Korea	capital area	contact with patient	6	#####	released
22	male	1973	Korea	Gwangju	contact with patient	16	#####	released
23	female	1962	China	capital area	visit to Wuhan		#####	released
24	male	1992	Korea	filtered at airport	residence in Wuhan		#####	released
25	female	1946	Korea	capital area	contact with patient	27	#####	isolated
26	male	1968	Korea	capital area	contact with patient	27	#####	isolated
27	female	1982	China	capital area	visit to China		#####	isolated
28	female	1989	China	capital area	contact with patient	3	#####	released
29	male	1938	Korea	capital area	contact with patient	83	#####	isolated

30	female	1952	Korea	capital area	contact with patient	29	#####	isolated
31	female	1959	Korea	Daegu			#####	isolated
32	female	2009	Korea	capital area	contact with patient	20	#####	released
33	female	1980	Korea	Daegu	contact with patient	31	#####	isolated
34	male	1996	Korea	Daegu			#####	isolated
35	female	1994	Korea	Daegu			#####	isolated
36	female	1972	Korea	Daegu			#####	isolated
37	male	1973	Korea	Gyeongsangbuk-do			#####	released
38	female	1963	Korea	Daegu			3/1/2020	deceased
39	female	1959	Korea	Gyeongsangbuk-do			3/2/2020	released
40	male	1943	Korea	capital area			3/3/2020	isolated
41	female	1951	Korea	Gyeongsangbuk-do			3/4/2020	isolated
42	female	1991	Korea	Daegu			3/5/2020	isolated
43	female	1962	Korea	Daegu			3/6/2020	isolated
44	female	1974	Korea	Daegu			3/7/2020	isolated
45	female	1966	Korea	Daegu			3/8/2020	isolated
46	male	1992	Korea	Daegu			3/9/2020	isolated
47	female	1957	Korea	Daegu	contact with patient	31	#####	isolated
48	female	1948	Korea	Daegu	contact with patient	31	#####	isolated
49	male	1962	Korea	Daegu	contact with patient	31	#####	isolated
50	male	1944	Korea	Daegu	contact with patient	31	#####	released
51	female	1959	Korea	Daegu	contact with patient	31	#####	released
52	male	1996	Korea				#####	isolated
53	female	1982	Korea	Daegu			#####	isolated
54	male	1963	Korea	Gyeongsangbuk-do			#####	deceased
55	male	1961	Korea	Gyeongsangbuk-do			#####	deceased
56	male	1945	Korea	capital area	contact with patient	83	#####	isolated
57	female	1968	Korea	Daegu			#####	released
58	female	1961	Korea				#####	isolated
59	female	1970	Korea	Daegu			#####	released
60	male	1998	Korea	Gyeongsangbuk-do	visit to Daegu		#####	isolated
61	female	1990	Korea	Gyeongsangbuk-do			#####	isolated
62	female	2000	Korea	Gyeongsangbuk-do			#####	isolated
63	male	1971	Korea				#####	isolated
64	female	1997	Korea				#####	isolated
65	male	1984	Korea				#####	isolated
66	female	1963	Korea				#####	isolated
67	male	1965	Korea				#####	isolated
68	female	1998	Korea	Gyeongsangbuk-do			#####	isolated
69	female	1986	Korea				4/1/2020	isolated
70	female	1955	Korea				4/2/2020	isolated
71	male	1956	Korea				4/3/2020	isolated
72	male	1944	Korea	capital area	contact with patient	6	4/4/2020	released

73	male	1951	Korea	Daegu			4/5/2020	isolated
74	female	1972	Korea	Gyeongsangbuk-do			4/6/2020	isolated
75	male	1989	Korea	Daegu			4/7/2020	isolated
76	female	1999	Korea	Daegu			4/8/2020	isolated
77	female	1979	Korea	Daegu			4/9/2020	isolated
78	female	1996	Korea	Daegu			#####	isolated
79	female	1979	Korea	Daegu			#####	isolated
80	female	1974	Korea	Gyeongsangbuk-do	visit to Daegu		#####	isolated
81	male	1957	Korea	Gyeongsangbuk-do			#####	deceased
82	female	1965	Korea	Gyeongsangbuk-do	visit to Daegu		#####	isolated
83	female	1991	Korea	Gyeongsangbuk-do	visit to Daegu		#####	isolated
84	male	1953	Korea	Gyeongsangbuk-do			#####	deceased
85	male	2000	Korea	capital area			#####	isolated
86	female	1941	Korea	capital area	contact with patient	136	#####	isolated
87	male	1992	Korea	Jeollabuk-do	visit to Daegu		#####	isolated
88	male	1946	Korea	Daegu			#####	deceased
89	female	1944	Korea	capital area	contact with patient	40	#####	isolated
90	male	1980	Korea	Gwangju	visit to Daegu		#####	isolated
91	male	1936	Korea	capital area	contact with patient	56	#####	isolated
92	male	1998	Korea	Jeju-do	visit to Daegu		#####	isolated
93	female	1988	Korea	Daegu	contact with patient	31	#####	isolated
94	male	1985	Korea	capital area			#####	isolated
95	male	1987	Korea	Daegu	contact with patient	31	#####	isolated
96	male	1989	Korea	Gwangju	contact with patient	126	#####	isolated
97	male	1961	Korea	capital area	visit to Daegu		#####	isolated
98	male	2001	Korea	Busan			#####	isolated
99	female	1965	Korea	Gyeongsangbuk-do			5/1/2020	deceased
100	female		Korea	Gyeongsangbuk-do	visit to Daegu		5/2/2020	isolated
101	male	1985	Korea	Gwangju	visit to Daegu		5/3/2020	isolated
102	female	1974	Korea	Gyeongsangbuk-do			5/4/2020	isolated
103	female	1998	Korea	Jeju-do	visit to Daegu		5/5/2020	isolated
104	male	1984	Korea	Jeollabuk-do	contact with patient	113	5/6/2020	isolated
105	female	1963	Korea	Busan			5/7/2020	isolated
106	female	1974	Korea	Gyeongsangbuk-do	visit to Daegu		5/8/2020	isolated
107	male	1968	Korea	Gyeongsangbuk-do			5/9/2020	isolated
108	male	1983	Korea	Gyeongsangbuk-do			#####	isolated
109	female	2000	Korea	Gyeongsangbuk-do	visit to Daegu		#####	isolated
110	male	1966	Korea	Gyeongsangbuk-do			#####	isolated
111	male	1985	Korea	Gwangju	visit to Daegu		#####	isolated
112	female	1978	Korea	Gyeongsangbuk-do	visit to Daegu		#####	isolated
113	female	1997	Korea	Daejeon	visit to Daegu		#####	isolated
114	female	1965	Korea	Gyeongsangbuk-do	visit to Daegu		#####	isolated
115	female	1946	Korea	Gyeongsangbuk-do	visit to Daegu		#####	isolated

116	female	1982	Korea	capital area	visit to Daegu		#####	isolated
117	male	1958	Korea	Gyeongsangbuk-do			#####	deceased
118	male	1962	Korea	Gyeongsangbuk-do			#####	deceased
119	male	1955	Korea	Gyeongsangbuk-do			#####	isolated
120	female	1943	Korea	Gyeongsangbuk-do			#####	isolated
121	female	1975	Korea	Gyeongsangbuk-do			#####	isolated
122	male	1984	Korea	Chungcheongbuk-do	contact with patient	230	#####	isolated
123	female	1985	Korea	Chungcheongbuk-do	contact with patient	230	#####	isolated
124	male		Korea	Chungcheongnam-do			#####	isolated
125	male	1996	Korea	capital area			#####	isolated
126	male	1980	Korea	capital area	pilgrimage to Israel		#####	isolated
127	male	1959	Korea	capital area			#####	isolated
128	male	1999	Korea	capital area	visit to Daegu		#####	isolated
129	male	1956	Korea	capital area	contact with patient	30	#####	isolated
130	female	1993	Korea	Ulsan	visit to Daegu		6/1/2020	isolated
131	male	1958	Korea	capital area	contact with patient	161	6/2/2020	isolated
132	male	1999	Korea	Gangwon-do	visit to Daegu		6/3/2020	isolated
133	female	1955	Korea	Daejeon			6/4/2020	isolated
134	female		Korea	Gangwon-do	visit to Daegu		6/5/2020	isolated
135	female		Korea	Gangwon-do	visit to Daegu		6/6/2020	isolated
136	male	1995	Korea	Gyeongsangbuk-do	visit to Daegu		6/7/2020	isolated
137	male	1956	Korea	Gyeongsangbuk-do	pilgrimage to Israel		6/8/2020	isolated
138	male	1956	Korea	Gyeongsangbuk-do	pilgrimage to Israel		6/9/2020	isolated
139	female	1960	Korea	Gyeongsangbuk-do	pilgrimage to Israel		#####	isolated
140	female	1960	Korea	Gyeongsangbuk-do	pilgrimage to Israel		#####	isolated
141	male	1966	Korea	capital area			#####	isolated
142	male	1987	Korea	Gyeongsangbuk-do			#####	isolated
143	female	1977	Korea	Gyeongsangbuk-do			#####	isolated
144	female	1996	Korea	Gyeongsangbuk-do			#####	isolated
145	female		Korea	Gangwon-do	visit to Daegu		#####	isolated
146	female	1960	Korea	capital area	visit to Daegu		#####	isolated
147	male		Korea	Gangwon-do			#####	isolated
148	male	1986	Korea	Daegu	visit to Daegu		#####	isolated
149	female	1954	Korea	capital area			#####	isolated
150	female	1957	Korea	Gyeongsangbuk-do	pilgrimage to Israel		#####	isolated
151	female	1959	Korea	capital area	contact with patient	246	#####	isolated
152	male	1979	Korea	Gyeongsangbuk-do			#####	deceased
153	female	1965	Korea	Gyeongsangbuk-do			#####	isolated
154	male		Korea	Gyeongsangbuk-do			#####	isolated
155	female	1968	Korea	capital area	contact with patient	347	#####	isolated
156	male	1985	Korea	Gwangju	contact with patient	126	#####	isolated
157	male	1974	Korea	capital area	visit to Daegu		#####	isolated
158	female	1962	Korea	Gyeongsangbuk-do			#####	isolated

159	female	1957	Korea	Gyeongsangbuk-do		#####	isolated
160	female	1977	Korea	Gyeongsangbuk-do		7/1/2020	isolated
161	male	1980	Korea	capital area	contact with patient	246	7/2/2020 isolated
162	female	1969	Korea	Gyeongsangbuk-do		7/3/2020	isolated
163	male	1957	Korea	Daejeon	contact with patient	372	7/4/2020 isolated
164	male	1997	Korea	capital area	visit to Daegu		7/5/2020 isolated
165	female	1985	Korea	Gwangju	contact with patient	126	7/6/2020 released
166	male		Korea	Gyeongsangbuk-do		7/7/2020	isolated
167	male	1968	Korea	capital area	visit to Vietnam		7/8/2020 isolated
168	male		Korea	Gyeongsangbuk-do		7/9/2020	isolated
169	female		Korea	Gyeongsangbuk-do	contact with patient	#####	isolated
170	female	1993	Korea	capital area		#####	isolated
171	male	1985	Korea	Gwangju	contact with patient	164	##### isolated
172	female	1997	Korea	Gyeongsangbuk-do		#####	isolated
173	male	1949	Korea	Daegu		#####	deceased
174	female	1958	Korea	Gyeongsangbuk-do		#####	isolated
175	male	1997	Korea	Gyeongsangbuk-do		#####	isolated
176	female	1950	Korea	capital area		#####	isolated

(۲)

این داده ها همان طور که بیان شده است مربوط به تعدادی از مبتلایان به بیماری کرونا در کشور کره جنوبی می باشد. تعداد این بیماران ۱۷۶ تا است و ۸ مشخصه برای هر یک از بیماران نوشته شده است. این مشخصه ها عبارتند از :

۱. جنسیت : جنسیت بیمار است.
 ۲. سال تولد : سال تولد بیمار است.
 ۳. کشور : کشوری است که بیمار در آن زندگی می کند.
 ۴. ناحیه : منطقه ای است که بیمار در آن زندگی می کند.
 ۵. دلیل بیماری : دلیلی است که باعث شده بیمار به کرونا مبتلا شود.
 ۶. افراد مبتلا شده : تعداد افرادی است که در اثر تماس با این بیمار ، کرونا گرفته اند.
 ۷. تاریخ تایید : تاریخی است که بیمار در بیمارستان بستری شده است.
 ۸. وضعیت : نشان می دهد که بیمار در چه وضعیتی (مثلا فوت شده ، مرخص شده) قرار دارد.
- در سطر های مختلف بیماران مختلف را می بینیم که هر کدام مقدار مشخصی برای مشخصه های بیان شده دارند. البته بعضی از بیماران تعدادی از مشخصه ها مانند افراد مبتلا شده و یا دلیل بیماری را ندارند.

اکثر بیماران از کشور کره جنوبی اند و سن شان در حدود ۶۰ سال است.

(۳)

فرمول استفاده شده برای بدست آوردن std به صورت زیر می باشد.

$$\sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}}$$

مقادیر به صورت زیر می باشند.

```
mean = 1973.5272727272727
max = 2009
std = 16.986208048239988
```

(۴)

با استفاده از متد isnull() از pandas متوجه می شویم که مقادیر null در جدول ما وجود دارند.

پس با استفاده از متد dropna() از pandas ، این مقادیر null را از جدول مان حذف می کنیم. که در نتیجه این کار جدول مان که قبلا ۱۷۶ سطر داشت ، اکنون ۴۲ سطر بیشتر ندارند.

(۵)

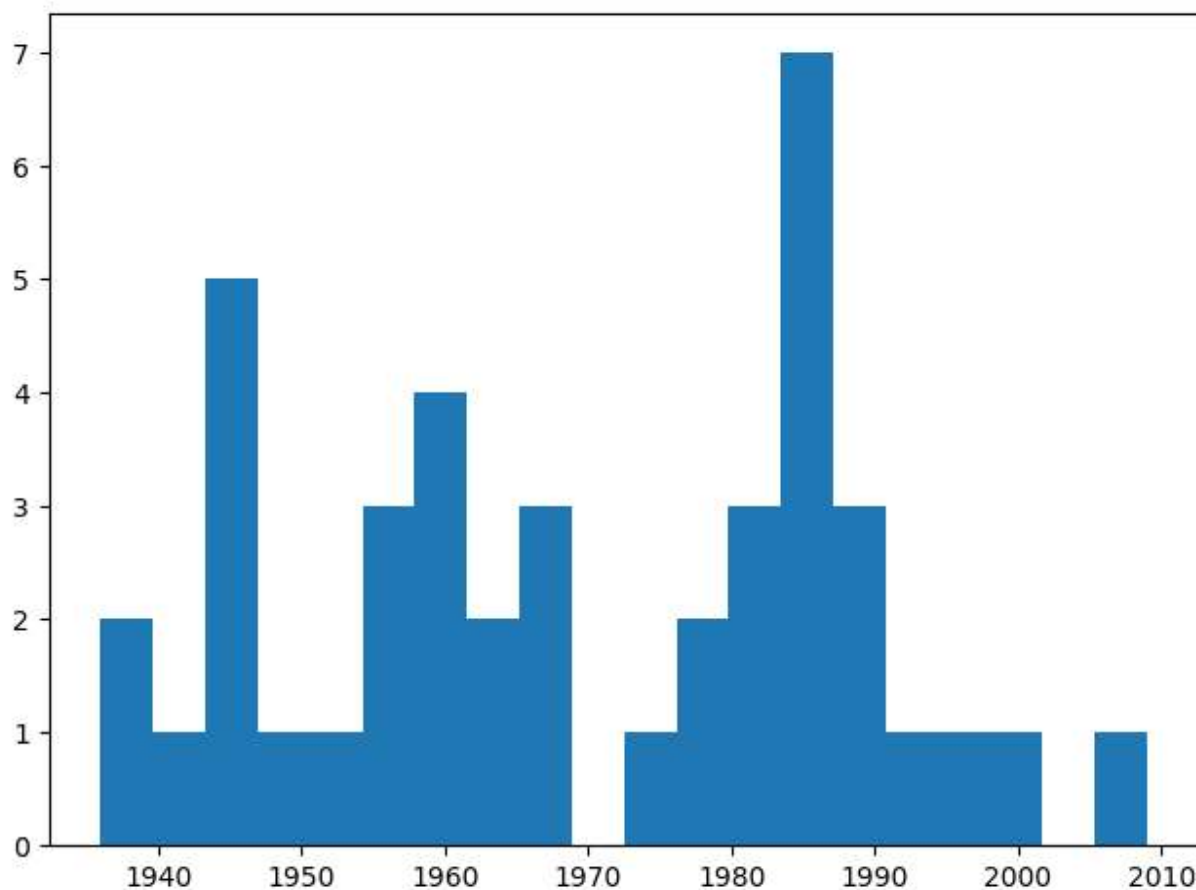
با توجه به اینکه در مرحله ۴ تمامی مقادیر null را حذف کرده ایم پس تمامی نمودار ها را با استفاده از ۴۱ داده باقی مانده رسم می کنیم.

histogram plot

برای رسم این نمودار از سال تولد افراد استفاده می کنیم به این صورت که 20 بازه را در نظر گرفته و تعداد افراد موجود در هر بخش را در نمودار نمایش می دهیم.

نمودار به صورت زیر می باشد.

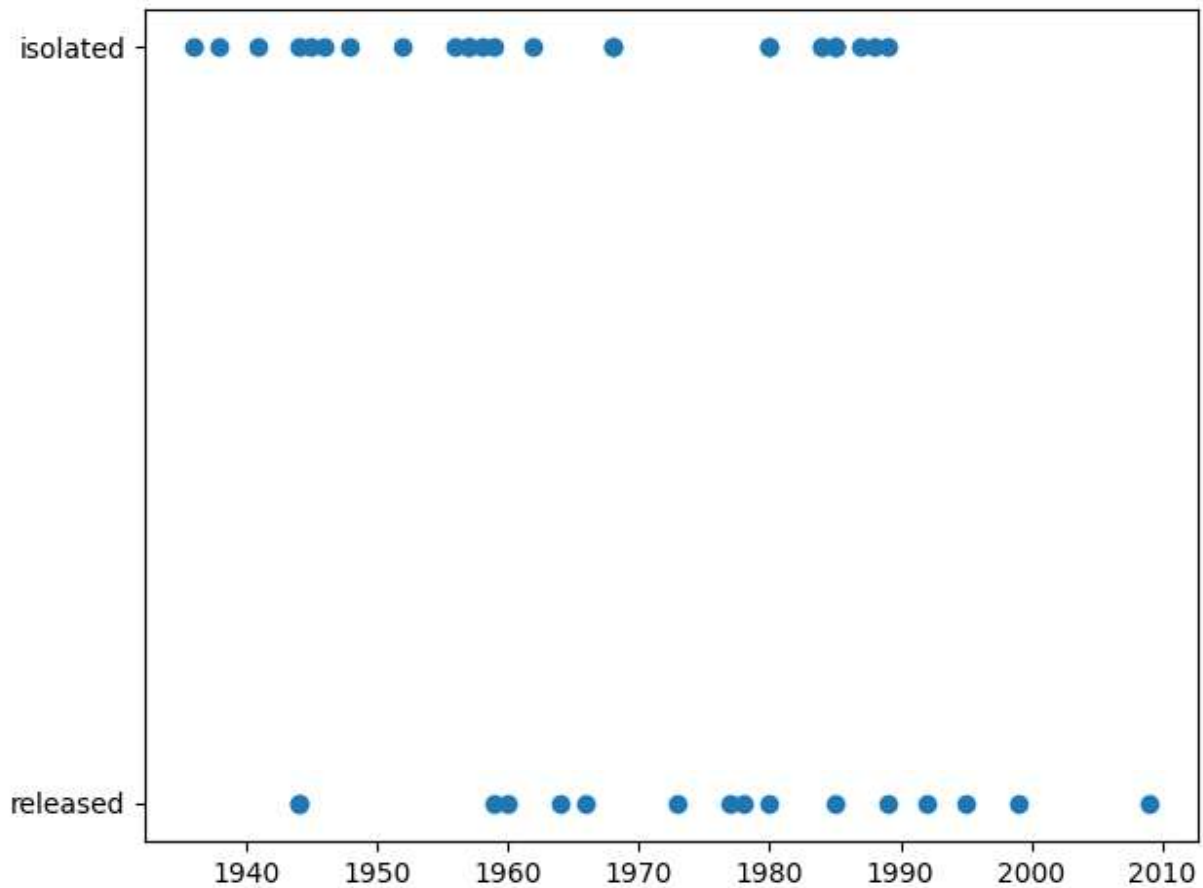
از این نمودار مشخص است که بیشتر افراد (۷ نفر) سال تولد شان بین 1984 تا 1988 قرار دارد.



Scatter plot

برای رسم این نمودار از سال تولد و وضعیت افراد استفاده می کنیم تا بتوانیم اطلاعات معنی داری بدست آوریم.
نمودار رسم شده در ادامه آمده است.

از این نمودار می توان به وضوح تشخیص داد که بیشتر افراد جوان تر مرخص شده اند و همچنین بیشتر افراد پیر تر در قرنطینه هستند.

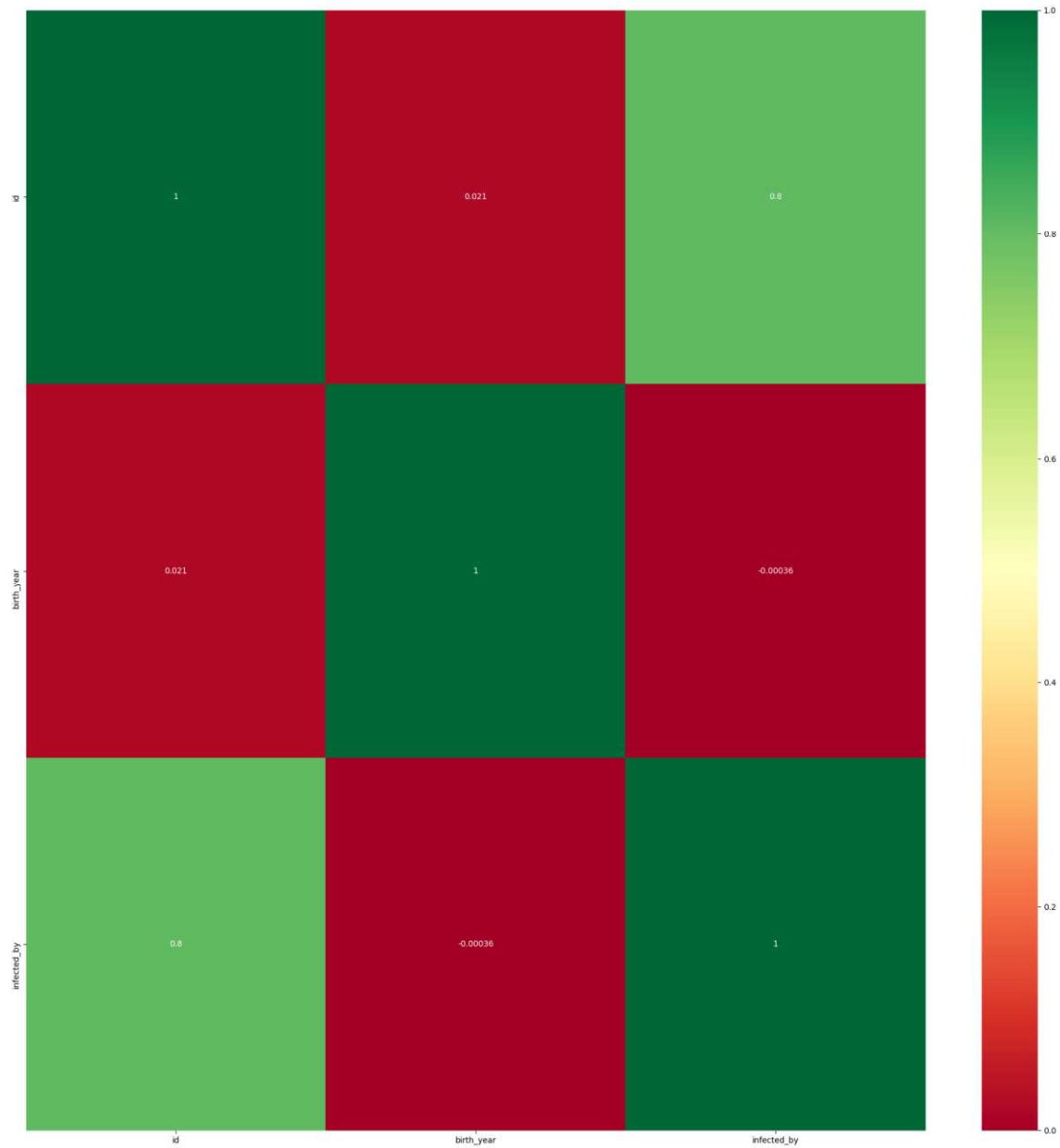


Matrix plot

برای رسم این نمودار از تمامی داده های عددی استفاده می کنیم. این داده ها عبارتند از شناسه ، سال تولد و افراد مبتلا شده. نمودار گفته شده به صورت زیر خواهد بود.

رابطه قوی بین شناسه و افراد مبتلا شده وجود دارد و این اتفاق به این معناست که افرادی که دیرتر در بیمارستان بستری شده اند نسبت به افراد ابتدایی ، افراد بیشتری را بیمار کرده اند.

Correlation در سایر قسمت ها منفی است. که به این معناست که هیچ رابطه مشخصی بین شان وجود ندارد.



(۶)

با بررسی نمودارهای رسم شده و همچنین داده های موجود ، outlier به نظر نمی آید که در این data set وجود داشته باشد.

اگر موجود بود می توانستیم یا آن outlier را متعادل کنیم و یا آن را حذف کنیم.

قسمت دوم :

با توجه به اینکه ۳۹۵ داده داریم و باید این داده ها را با نسبت ۸۰ به ۲۰ به دو قسمت **train** و **test** تقسیم کنیم، ۳۲۰ داده ابتدایی را برای **train** در نظر گرفته و ۷۵ داده باقی مانده را نیز برای **test** استفاده می کنیم. آموزش :

ابتدا مقادیر تصادفی بین ۰ و ۱ در لیست **A** که همان ضرایب مان می باشد وارد می کنیم. سپس هزینه (تابع کمترین مربعات) را با توجه به این **A** بدست می آوریم. سپس در آن نقطه گرادیان تابع هزینه نسبت به ضرایب بدست می آوریم. و در انتها مقدار قبلی ضرایب **A** را از حاصل ضرب نرخ یادگیری در گرادیان بدست آمده در مرحله قبل کم می کنیم.

با چاپ کردن مقدار تابع **cost** در طی آموزش متوجه می شویم که مقدار اولیه آن در حدود ۳۰۰,۰۰۰ است که پس از آموزش مقدار آن به حدود ۱۰۰۰ می رسد.

تست :

با توجه به ضرایب موجود در لیست **A** مقدار مقادیر موجود در هر یک از داده هر را در این ضرایب ضرب کرده و مقدار نهایی آن را که توسط مدل پیش بینی شده است بدست می آوریم. برای بدست آوردن دقت کار چهار حالت در نظر می گیریم.

۱. در این حالت تمامی داده های تست بررسی شده اند و پیش بینی هایی درست در نظر گرفته شده اند که اختلاف بین نمره واقعی و نمره پیش بینی شده کمتر از ۱ باشد.
۲. در این حالت تمامی داده های تست بررسی شده اند و پیش بینی هایی درست در نظر گرفته شده اند که اختلاف بین نمره واقعی و نمره پیش بینی شده کمتر از ۲ باشد.
۳. در این حالت داده هایی از مجموعه تست بررسی شده اند که نمره نهایی شان مخالف ۰ باشد و پیش بینی هایی درست در نظر گرفته شده اند که اختلاف بین نمره واقعی و نمره پیش بینی شده کمتر از ۱ باشد.
۴. در این حالت داده هایی از مجموعه تست بررسی شده اند که نمره نهایی شان مخالف ۰ باشد و پیش بینی هایی درست در نظر گرفته شده اند که اختلاف بین نمره واقعی و نمره پیش بینی شده کمتر از ۲ باشد.

علت در نظر گرفتن دو حالت ۳ و ۴ در این است که برخی از دانش آموزان که انتظار می رفت نمره ای در حد ۶ و ۷ کسب کنند ۰ کسب کرده اند و این داده ها که نوعی داده پرت به حساب می آیند باعث کاهش دقت مدل شده بودند.

نمونه هایی از خروجی کد که به ترتیب شامل ۴ نوع دقت بیان شده است در زیر آمده است.

```
accuracy(delta < 1) : 0.6578947368421053
accuracy(delta < 2) : 0.8289473684210527
accuracy(delta < 1) (no 0) : 0.7424242424242424
accuracy(delta < 2) (no 0) : 0.9393939393939394

accuracy(delta < 1) : 0.5394736842105263
accuracy(delta < 2) : 0.8289473684210527
accuracy(delta < 1) (no 0) : 0.6060606060606061
accuracy(delta < 2) (no 0) : 0.9242424242424242

accuracy(delta < 1) : 0.5921052631578947
accuracy(delta < 2) : 0.7631578947368421
accuracy(delta < 1) (no 0) : 0.6818181818181818
accuracy(delta < 2) (no 0) : 0.8636363636363636

accuracy(delta < 1) : 0.618421052631579
accuracy(delta < 2) : 0.7894736842105263
accuracy(delta < 1) (no 0) : 0.7121212121212122
accuracy(delta < 2) (no 0) : 0.8939393939393939
```