Tokyo 2020 Olympics and COVID-19:
An Exploration of the Influence of the Sports Event on Infections and Social Gatherings

Davide Talevi - 0001043409
davide.talevi@studio.unibo.it

Human Data Science a.a. 2022-2023

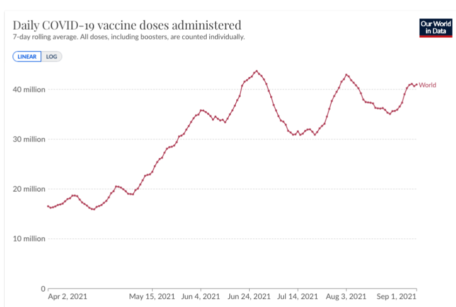
Abstract

The Tokyo 2020 Olympics represented an extraordinary event that faced unprecedented challenges due to the COVID-19 pandemic. Despite the pandemic's lasting impact on social gatherings and celebrations, the huge economic appeal of major sporting events has remained evident. In particular, this study will focus on the edition of the Tokyo 2020 Olympics, held in 2021, a moment of great joy and competition for athletes and spectators around the world. Despite the challenges posed by the pandemic, these Olympics demonstrated humanity's resilience and determination in pursuing the passion for sport and the sense of global unity that only events of this magnitude can generate. The 2021 edition of the Tokyo Olympics catalysed the participation of 3 billion viewers around the world, whether through television broadcasts, online platforms and celebrations in public and private places, Stating that 3 out of 4 people have followed the races[1]. This study intends to examine the effect of the Olympics on the increase in COVID-19 cases in different regions of the world, highlighting the potential impact of large meetings and celebrations related to these sporting events. To understand how events of this magnitude can affect the spread of infectious diseases such as COVID-19, this study proposes two methods of analysis based on the data collected. The first methodology involves the use of a Bayesian regression to multiple segments, able to analyse the data with flexibility and precision. Our analysis considers the impact of social gatherings in public places such as pubs, bars and stadiums. Using specific Bayesian models, we identified the periods when these social activities intensified and assessed whether there was a relationship between such increases and the increase in infections. The second methodology adopts a prediction model called ARIMA[2] that takes into account the vaccination progress of the respective countries studied, dividing the time period into arbitrary segments according to the indications provided by the medical literature. Using this model we will go to see if the trend of infections in the period of the Olympic Games followed the trend of the past months or we find a gap with the prediction to suggest that the population despite the commitment to vaccinations has had more contact and therefore an increase in infections. Through these analyses, we aim to provide a clearer and more in-depth perspective on the impact of the Tokyo 2020 Olympics on the spread of COVID-19, providing useful information to address and manage future sporting events of this magnitude safely and responsibly.

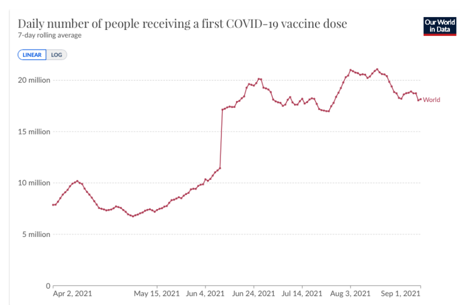
1 Introduction

The Tokyo 2020 Olympics was an event of great importance, bringing together athletes from all over the world to compete in different sports disciplines. These games caught the attention of a wide audience, with millions of spectators joining together to follow competitions live, both in sports

arenas and from home. However, the organisation of major sporting events during the COVID-19 pandemic has raised concerns regarding the spread of the virus.
It is therefore natural for us to ask ourselves questions, what is the evolution when such aggregations become the vehicle through which a life-threatening disease spreads? Is there a relationship between this scenario and, consequently, the wide range of small meetings that have emerged from it, and a possible variation in the trend of new daily cases? The article by L.Casini and M.Roccetti [3] explores the similar situation for the 2020 European football championships, let's analyse in this study the edition of the 2020 Olympics played in 2021 to have a more complete history of cases. The objective of this study is to examine whether there are correlations between the Olympics and related social gatherings and a possible change in the trend of new cases. These Olympics took place in a context where COVID-19 was already a known threat, so organisers took strict measures to ensure the safety of athletes and the public. Health restrictions and protocols have been implemented to mitigate the risk of virus transmission during the event. In particular, the competition took place with a focus on social distance and hygiene guidelines. This study focusses on the 2021 Olympics edition to get a comprehensive overview of COVID-19 cases during the event. Data on new cases in different states of the world are analysed to understand if there is a correlation between the rallies resulting from the Olympics and a potential increase in new cases. It is crucial to point out that within this analysis considerable weight was attributed to the intrinsic effect of the immunisation process underway during 2021. During this time period, an impressive vaccination effort was conducted with unprecedented scope, resulting in the daily inoculation of millions of doses of vaccine against the SARS-CoV-2 virus. This phenomenon may have caused a substantial influence on the epidemiological dynamics taken into account during the time interval of interest. There is also an important observation that many jurisdictions and sovereign states have issued vaccination mandates in order to achieve an adequate level of herd immunity and contain the spread of the virus. A striking example has emerged in Europe, where the 'Green Pass'[4] was introduced which established the mandatory nature of vaccination certificates in order to be able to access certain public places and participate in events of social relevance. This coercive approach has triggered an interesting dynamic of privacy considerations and the balance between individual freedom and collective security, confirming the breadth of the impact of vaccination strategies on contemporary society.



(a) Daily adminitered vaccine in the world.



(b) Daily people that receive the first vaccine dose in the world.

Figure 1: Trend of the world's vaccinations from Our World in Data resources[5].

The rest of the document is structured as follows: In the next section, we describe more precisely the data we used, their sources and the methodologies we used. Section 3 presents the results we have obtained, while Section 4 discusses them, along with their limitations, and concludes the document, presenting our final considerations.

2 Data Collection and Analysis

This section explains how data was collected for the project and what methods were applied to analyse it and collect the results.

The main focus of this study is the number of COVID-19 cases in each state during the period of the 2021 edition of the Tokyo Olympic Games, which was held on July 23, 2021 with the opening ceremony and ended on August 8, 2021. The data was retrieved from several sources. Data related to OWID[5] and WHO[6] infections, this amount of data is obviously excessive, so the dataset has been compressed and reduced as follows:

- The dataset mainly used for better data cleaning was that of the World Health Organisation, but information regarding the population of the respective countries was taken from the others[5].
- The iso_codes for whole continents have been removed, leaving only those of the countries.
- The time frame has been greatly shortened: this study considers the data from July 8, 2021 to August 15, 2021, from two weeks before the first race (to understand the underlying trend) to two weeks after the last (to take into account the incubation time and see the following trends)[7].
- One hypothesis to explain the results that will be shown in a later paragraph is that the regulations in place at the state level have facilitated the spread of the virus. As regulations are difficult to collect as pure data has been omitted from the study.
- To make the values of the new daily cases homogeneous and eliminate the jumps due to holidays and weekdays when the offices did not publish today's data, an additional column was created with the moving average with a 7-day window for both new infections and vaccinations. When we talk about values from here on, we mean this difference in moveable averages.
- In particular, the analysis will focus on 2 phases: one with the exchange points identified using a Bayesian technique of exchange points and another where the time series is divided into 4 time periods, namely before the first race, a week later, a week after the last race where a prediction model will be performed on each interval. This will shed light on the effects of a sporting event of this size, immediate and prolonged according to two separate analysis techniques.

While for the dataset of daily vaccines we took a public dataset on Kaggle[8] where we find the gionalieri data with the number of doses and boosters administered. The same manipulation methods were applied as the previous dataset. In addition, columns were renamed to work with the same keys and the iso_codes were transformed from ISO 3166-1 alpha-2 to ISO 3166-1 alpha-3 format[9].

For each of the two studies covered, the same study method is then performed for each country within the dataset, going to cycle for each iso_code present you go to apply the study method. It is important to emphasise that the purpose of these analyses is not to calculate with extreme precision the spread of the virus but only to quantify the moment in time when a sharp change in trend has occurred and to define these trends as increasing or decreasing to see the phenomenon on a general scale. The second method consists of separating the time series into 4 segments and using an ARIMA model previously trained on the data of the respective country taking into account the trend of the vaccination curve from the first dose administered in that country to calculate the trend for each section, complete with its slope coefficient. This method is medicine-driven because several articles like [10] show that COVID-19 incubates up to a week in patients before it manifests, so we can expect a spike in the second week of the time series. The other intervals are necessary to understand the underlying trends and whether the expected peak is prolonged over time or if it is just a special event. The reader must keep in mind, to understand the results, that in both techniques a negative slope means that the phenomenon is decreasing and a positive slope means that the phenomenon tends to increase therefore, since the phenomenon in question are new cases of a deadly disease, a the negative slope is desirable and the more negative, the better.

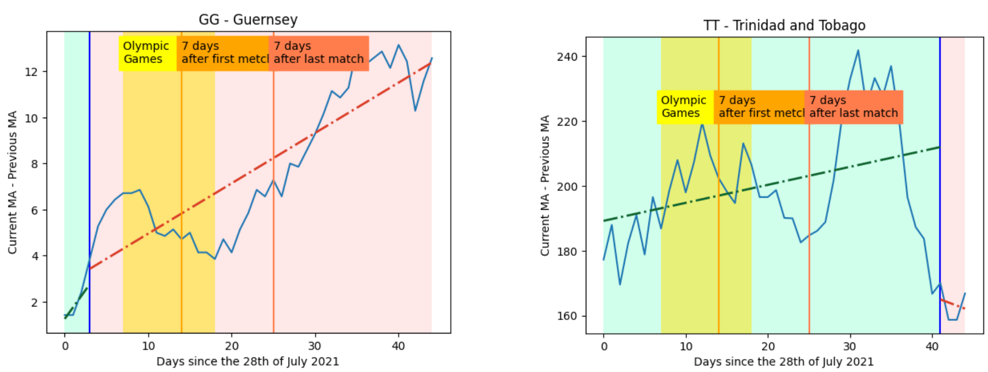
This analysis was performed on a Jupyter notebook in Python 3.10.9 using the matplotlib library to track the results, pymc3 for the analysis of Bayesian change points and the statsmodels library for ARIMA models functions.

3 Results

This section will explore the results of the two techniques used to understand the trends in this study: the Bayesian detection of the exchange point and the ARIMA prediction model to see the expected trend of new cases with respect to the vaccination trend. This section should contain 412 images: 206 for each state but, for brevity, all images are available for viewing at a drive folder linked in the appendix after the conclusions of this article, and this section will present only a few interesting examples, along with the summary tables of the data found.

3.1 Bayesian Changepoint detection

The initial section focusses on the Bayesian methodology of turning point detection, used to identify the phases in which variations in time series models occurred. The selected algorithmic approach has been configured to detect the initial point of change in the data, with a customised sensitivity for each state condition. The exchange point found was not always the same day for all states but, on average, the exchange point was 20.89 days (95% CI: 6 — 7.4)[10] after the start of the study (the date of the first game played is 7 days after the start). The minimum exchange point was found in Guernsey on day 3 and the maximum exchange point was found in Trinidad and Tobago on day 41 of the 44 on which the study focusses. The Guernsey chart in particular shows an interesting trend of these charts, namely that after the trading point the trend will not reverse but its slope will decrease by about half, which means a greater and constant spread of the virus.

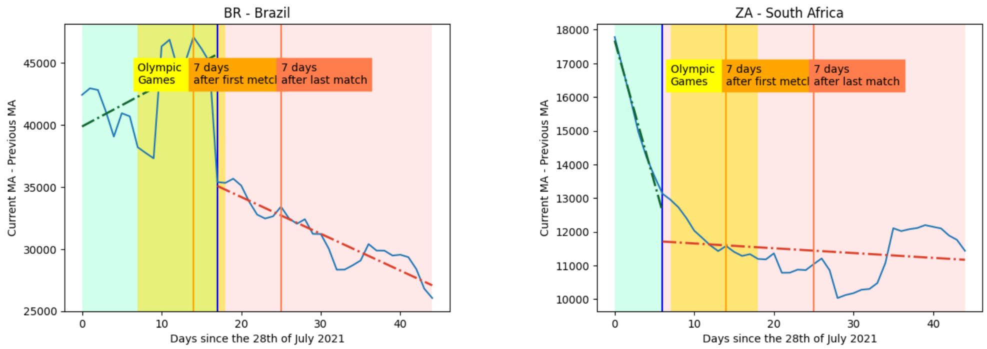


(a) Guernsey currency exchange point chart.

(b) Trinidad e Tobago's exchange point chart.

Figure 2: Charts showing the minimum and maximum of chanpoints found.

Altogether we have 129 countries that have had an increase in infections after the point of exchange while 77 countries have decreased. Only 56 out of 206 states experienced a trend reversal, of which 25 had a worsening of cases and 31 an improvement, hence an increase in cases. The country with the greatest deterioration was Brazil with -63924,87% and the one with the greatest positive trend was South Africa with 83118,76%. The total weighted average of the percentage change after trading points is 11568,52%. This means that, despite having a strong positive vaccination trend and with more defined restrictions (e.g. EU Green Pass), after the exchange point there is a considerable increase in daily cases. Although interesting, the data reported from all countries do not clearly link the inclinations of the cases to the Olympic competitions, in fact although many, not all the countries in our study had the opportunity to participate. This study therefore places an interest in deepening a part of these states.



(a) Brazil currency exchange point chart.

(b) South Africa's exchange point chart.

Figure 3: Graphs showing the minimum and maximum increase percentage after the found chanpoints.

Table 1 shows, for all states, their name, the value of the slope before the exchange point, the value of the slope after the exchange point, and when the exchange point occurred. Only for 2 countries a second exchange point was also found in the period studied.

	State	Iso_code	FirstSlope	SecondSlope	when	ThirdSlope
0	Andorra	AD	-0.229402	-0.566917	26	nan
1	United Arab Emirates	AE	-3.28358	-20.2016	31	nan
2	Afghanistan	AF	-23.5023	-2.42857	36	nan
3	Antigua and Barbuda	AG	0.139255	0.899443	24	nan
4	Anguilla	AI	-0.00504202	0.297269	16	nan
5	Albania	AL	1.43395	26.7632	21	nan
6	Armenia	AM	5.10963	13.9881	27	nan
7	Angola	AO	-0.0533886	4.13766	34	nan
8	Argentina	AR	-155.42	-202.146	30	nan
9	Austria	AT	11.3028	44.744	28	nan
10	Australia	AU	9.10145	40.6429	30	nan
11	Aruba	AW	1.03047	2.71677	12	nan
12	Azerbaijan	AZ	14.3132	124.961	13	nan
13	Bosnia and Herzegovina	BA	0.405808	13.6796	13	nan
14	Barbados	BB	-0.197336	0.453546	33	nan
15	Bangladesh	BD	22.9776	-332.259	32	nan
16	Belgium	BE	16.9571	14.8587	20	nan
17	Burkina Faso	BF	-0.121193	0.0748848	14	nan
18	Bulgaria	BG	3.3	41.2266	16	nan
19	Bahrain	BH	-0.835165	-0.159824	13	nan
20	Burundi	BI	1.95408	2.62197	7	nan
21	Benin	BJ	0	3	nan	20.0165
22	Saint Barthélemy	BL	0.0193878	-1.12217	15	nan
23	Bermuda	BM	0.0180534	0.343842	13	nan
24	Brunei Darussalam	BN	-0.109752	9.13117	24	nan
25	Bolivia (Plurinational State of)	BO	-26.2054	-4.76683	21	nan
26	Brazil	BR	342.59	-296.659	17	nan
27	Bahamas	BS	3.94006	1.01194	12	nan
28	Bhutan	BT	-0.483306	-0.0690476	36	nan
29	Botswana	BW	15.9957	-49.4506	34	nan

	State	Iso_code	FirstSlope	SecondSlope	when	ThirdSlope
30	Belarus	BY	3.67439	26.4693	35	nan
31	Belize	BZ	0.621534	2.36176	29	nan
32	Canada	CA	5.3699	81.1283	15	nan
33	Democratic Republic of the Congo	CD	-1.6379	-5.47441	31	nan
34	Central African Republic	CF	0.0881875	0.115884	33	nan
35	Congo	CG	-0.319481	0.132315	11	nan
36	Switzerland	CH	19.9101	80.2235	25	nan
37	Côte d'Ivoire	CI	2.02696	3.12851	17	nan
38	Chile	CL	-57.6792	-19.4473	19	nan
39	Cameroon	CM	0.105006	4.33896	27	nan
40	China	CN	2.00194	-2.65065	34	nan
41	Colombia	CO	-544.86	-150.713	30	nan
42	Costa Rica	CR	1.07082	25.9696	27	nan
43	Cuba	CU	79.4401	29.781	12	nan
44	Cabo Verde	CV	-0.335802	0.867857	30	nan
45	Curaçao	CW	4.05714	-0.302506	5	nan
46	Cyprus	CY	-14.5355	-13.5433	18	nan
47	Czechia	CZ	-3.29004	0.269068	10	nan
48	Germany	DE	74.1375	292.693	30	nan
49	Djibouti	DJ	0.0115724	0.0322129	28	nan
50	Dominica	DM	0.0159627	1.63766	24	nan
51	Dominican Republic	DO	-36.151	-4.24511	6	nan
52	Algeria	DZ	7.22857	-29.7212	31	nan
53	Ecuador	EC	-16.1353	-3.61729	26	nan
54	Estonia	EE	4.00966	5.55271	16	nan
55	Egypt	EG	-7.70408	3.66677	7	nan
56	Eritrea	ER	-0.743043	-0.0670807	21	nan
57	Spain	ES	33.5165	-461.936	24	nan
58	Ethiopia	ET	2.95212	33.689	13	nan
59	Finland	FI	12.398	1.00772	13	nan
60	Fiji	FJ	1.28632	-11.5918	27	nan
61	Faroe Islands	FO	-0.573469	-0.0516447	15	nan
62	France	FR	424.327	75.9471	10	nan
63	Gabon	GA	-0.156201	0.374023	27	nan
64	The United Kingdom	GB	162.249	187.921	11	nan
65	Grenada	GD	0.0164596	0.37384	24	nan
66	Georgia	GE	92.8159	54.4397	20	nan
67	French Guiana	GF	0.89761	3.11573	27	nan
68	Guernsey	GG	0.5	0.218274	3	nan
69	Ghana	GH	0.243287	-3.28308	20	nan
70	Gibraltar	GI	-0.00327348	-0.529082	30	nan
71	Greenland	GL	0.0629991	-0.0585827	22	nan
72	Gambia	GM	-0.261465	-1.7	34	nan
73	Guinea	GN	-0.80314	-1.12085	14	nan
74	Guadeloupe	GP	2.17602	25.5589	15	nan
75	Equatorial Guinea	GQ	-0.0123759	1.90779	34	nan
76	Greece	GR	19.8207	7.91071	30	nan
77	Guatemala	GT	14.8774	50.4002	22	nan
78	Guam	GU	-0.0442282	3.39377	18	nan
79	Guinea-Bissau	GW	0.995504	0.478848	12	nan

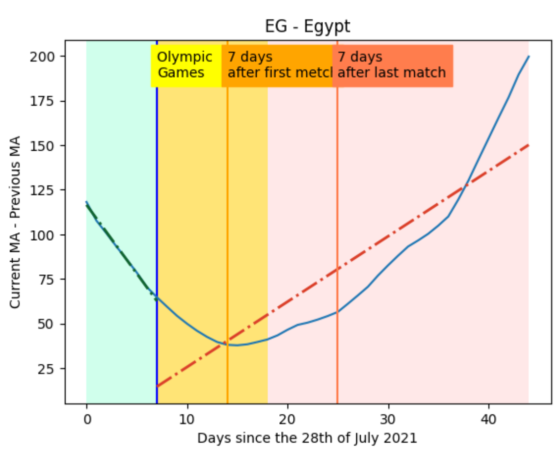
	State	Iso_code	FirstSlope	SecondSlope	when	ThirdSlope
80	Guyana	GY	-0.497679	5.09784	35	nan
81	Honduras	HN	21.9466	3.68852	13	nan
82	Croatia	HR	3.72723	16.1473	29	nan
83	Haiti	HT	-0.470232	-0.515038	26	nan
84	Hungary	HU	0.52356	4.64396	31	nan
85	Indonesia	ID	-666.283	-1000.29	28	nan
86	Ireland	IE	28.1077	18.3842	25	nan
87	Israel	IL	117.334	254.089	26	nan
88	Isle of Man	IM	-1.07079	-1.58807	31	nan
89	India	IN	-79.8477	654.689	38	nan
90	Iraq	IQ	33.0276	-186.257	36	nan
91	Iran (Islamic Republic of)	IR	482.257	222.202	15	nan
92	Iceland	IS	1.58571	0.439897	9	nan
93	Italy	IT	263.405	59.3483	11	nan
94	Jersey	JE	-9.84373	-0.237291	24	nan
95	Jamaica	JM	7.82274	20.6313	28	nan
96	Jordan	JO	-8.52198	0.303912	13	nan
97	Japan	JP	191.998	570.638	13	nan
98	Kenya	KE	15.762	9.76434	14	nan
99	Kyrgyzstan	KG	-25.3625	-15.4769	31	nan
100	Cambodia	KH	-12.5792	-5.41813	20	nan
101	Comoros	KM	-0.142857	0.0519587	17	nan
102	Saint Kitts and Nevis	KN	0.0239055	2.38242	31	nan
103	Republic of Korea	KR	5.91708	0.86475	24	nan
104	Kuwait	KW	-33.4795	-19.726	34	nan
105	Cayman Islands	KY	0.0882353	0.0448041	17	nan
106	Kazakhstan	KZ	170.955	-7.36383	12	nan
107	Lao People's Democratic Republic	LA	7.57313	4.38004	8	nan
108	Lebanon	LB	21.1436	15.8523	13	nan
109	Saint Lucia	LC	-0.317577	4.99187	17	nan
110	Liechtenstein	LI	-0.00032967	0.142965	25	nan
111	Sri Lanka	LK	52.6418	133.392	25	nan
112	Liberia	LR	-2.92381	-0.0358154	9	nan
113	Lesotho	LS	-0.000367715	-0.438095	36	nan
114	Lithuania	LT	6.89116	13.3408	8	nan
115	Luxembourg	LU	-2.6348	-0.254789	17	nan
116	Latvia	LV	1.52372	3.81071	30	nan
117	Libya	LY	-4.19398	-80.5265	39	nan
118	Morocco	MA	147.704	14.804	14	nan
119	Monaco	MC	0.420408	-0.210237	6	nan
120	Republic of Moldova	MD	4.37973	17.952	33	nan
121	Montenegro	ME	1.02857	18.0362	9	nan
122	Saint Martin	MF	-0.0163644	0.131258	18	nan
123	Madagascar	MG	-0.348794	-0.112919	21	nan
124	North Macedonia	MK	2.34105	47.5071	23	nan
125	Mali	ML	0.00903444	0.129164	22	nan
126	Myanmar	MM	-35.4494	-77.5281	22	nan
127	Mongolia	MN	-7.29989	86.5048	36	nan
128	Northern Mariana Islands (Commonwealth of the)	MP	0.00105042	0.00183752	17	nan
129	Martinique	MQ	15.113	-13.5683	13	nan

	State	Iso_code	FirstSlope	SecondSlope	when	ThirdSlope
130	Mauritania	MR	5.0314	-1.84329	13	nan
131	Malta	MT	2.53186	-1.01916	17	nan
132	Mauritius	MU	1.42231	11.8117	34	nan
133	Malawi	MW	-4.95394	-15.2125	28	nan
134	Mexico	MX	463.596	24.6302	10	nan
135	Malaysia	MY	337.412	198.188	13	nan
136	Mozambique	MZ	-1.91795	-41.8742	29	nan
137	Niger	NE	-0.208315	0.0152625	18	nan
138	Nigeria	NG	13.2423	8.90733	15	nan
139	Nicaragua	NI	0.292696	1.94286	40	nan
140	Netherlands	NL	-229.983	-32.6877	17	nan
141	Norway	NO	5.91837	20.6982	13	nan
142	Nepal	NP	25.6857	-11.3701	11	nan
143	New Zealand	NZ	-0.0229199	1.47091	14	nan
144	Oman	OM	-67.2211	-10.5042	8	nan
145	Panama	PA	-7.8697	-14.5425	33	nan
146	Peru	PE	-16.5619	-12.8603	29	nan
147	French Polynesia	PF	14.2557	7.83588	25	nan
148	Papua New Guinea	PG	-0.46116	0.654433	22	nan
149	Philippines	PH	130.79	372.853	22	nan
150	Pakistan	PK	42.7214	-16.244	15	nan
151	Poland	PL	2.60977	3.14305	19	nan
152	Puerto Rico	PR	22.8017	4.99634	18	nan
153	occupied Palestinian territory, including east Jerusalem	PS	3.75266	72.1794	27	nan
154	Portugal	PT	13.375	-12.931	17	nan
155	Paraguay	PY	-30.5967	-8.26274	33	nan
156	Qatar	QA	2.42379	2.47904	24	nan
157	Réunion	RE	-2.33766	-0.508756	12	nan
158	Romania	RO	3.06407	17.0138	10	nan
159	Serbia	RS	16.8419	83.5011	29	nan
160	Rwanda	RW	-12.6696	-10.2855	24	nan
161	Saudi Arabia	SA	-8.26413	-31.8144	28	nan
162	Seychelles	SC	0.216013	0.380876	14	nan
163	Sudan	SD	-0.767161	-0.0244099	21	nan
164	Sweden	SE	15.4286	16.6664	19	nan
165	Singapore	SG	9.32857	-2.26595	4	nan
166	Slovenia	SI	2.69188	13.7286	30	nan
167	Slovakia	SK	1.01499	1.56727	12	nan
168	Sierra Leone	SL	-1.18988	-0.148363	22	nan
169	San Marino	SM	0.110654	0.7	40	nan
170	Senegal	SN	-8.62651	-15.9786	36	nan
171	Somalia	SO	1.19903	0.526484	20	nan
172	Suriname	SR	-2.9619	1.47029	9	nan
173	South Sudan	SS	1.69152	0.41139	13	nan
174	Sao Tome and Principe	ST	0.00643501	0.769048	36	nan
175	El Salvador	SV	-10.0597	0.233293	10	nan
176	Sint Maarten	SX	0.581245	0.246617	26	nan
177	Syrian Arab Republic	SY	-0.263736	3.02197	13	nan
178	Eswatini	SZ	3.78528	7.409	10	nan
179	Turks and Caicos Islands	TC	0.0163749	0.0305719	23	nan

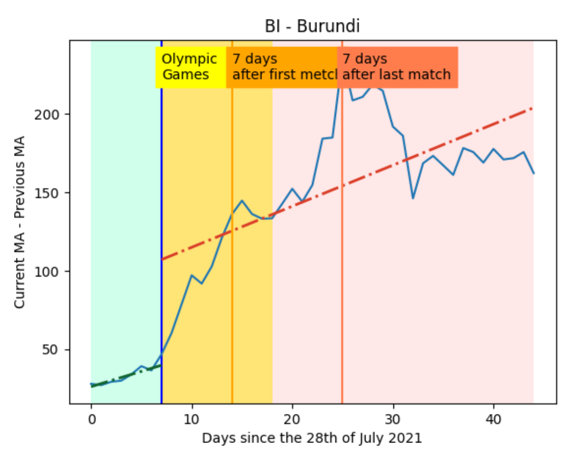
	State	Iso_code	FirstSlope	SecondSlope	when	ThirdSlope
180	Chad	TD	0.120748	-0.00745106	8	nan
181	Togo	TG	2.94317	5.3712	14	nan
182	Thailand	TH	482.38	99.7647	14	nan
183	Tajikistan	TJ	1.35034	-0.826085	8	nan
184	Timor-Leste	TL	-1.71978	8.61489	13	nan
185	Tunisia	TN	-290.209	-15.0607	18	nan
186	Türkiye	TR	427.254	21.2487	13	nan
187	Trinidad and Tobago	TT	0.555326	-0.942857	41	nan
188	United Republic of Tanzania	TZ	-1	262	nan	-8.81366
189	Ukraine	UA	10.7349	29.3399	23	nan
190	Uganda	UG	-8.15501	-190.797	34	nan
191	United States of America	US	2151.8	2825.04	17	nan
192	Uruguay	UY	-13.9689	-1.40909	22	nan
193	Uzbekistan	UZ	3.69091	4.30196	10	nan
194	Saint Vincent and the Grenadines	VC	0.0361781	0.0349068	21	nan
195	Venezuela (Bolivarian Republic of)	VE	7.79307	-2.81869	10	nan
196	British Virgin Islands	VG	-6.0993	-0.2	22	nan
197	United States Virgin Islands	VI	0.573529	0.132794	16	nan
198	Viet Nam	VN	413.252	128.761	10	nan
199	Bonaire	XA	0.0305414	0.0571429	40	nan
200	Kosovo[1]	XK	0.00952381	66.8144	9	nan
201	Yemen	YE	-0.0367347	1.09747	15	nan
202	Mayotte	YT	-0.136578	0.374885	14	nan
203	South Africa	ZA	-845.441	-14.2532	6	nan
204	Zambia	ZM	-40.7899	-5.25255	30	nan
205	Zimbabwe	ZW	-54.679	-23.7203	27	nan

Table 1: Table showing all information on technical Bayesian exchange points.

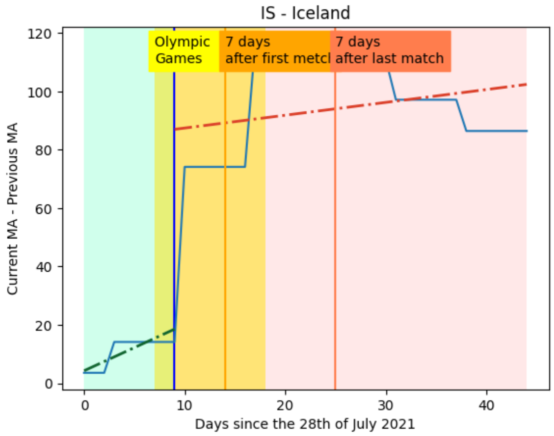
Now let's see some examples of interesting cases for this technique that shows ideal cases to demonstrate the hypothesis that the Olympic Games have influenced the trend of new cases: in Egypt the exchange point is exactly the date of the first Olympic race and a clear change in trend is evident with an increase of 1137.08% (slope from -7.704 to 3.667) compared to the trend We find the same trend in Burundi with a trend change of 66.79% (slope from 1,954 to 2,622). This situation is not shared in the 'virtuous' states, such as Italy, France, the United Kingdom, Romania, Japan and Iceland where the trend continues to rise but with the exchange point a few days later than the start of the races (3-7 days later).



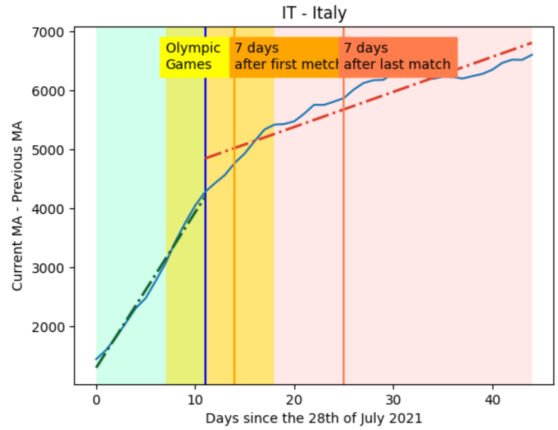
(a) Egypt's exchange rate chart.



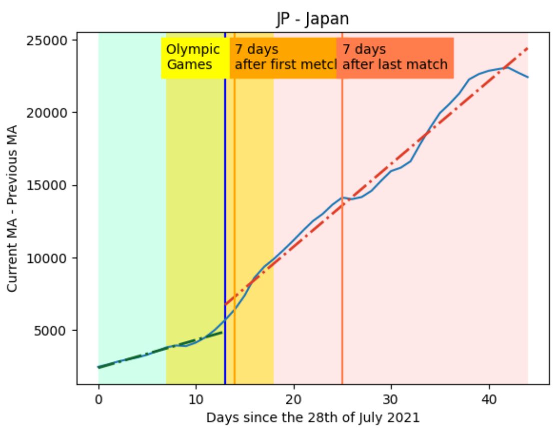
(b) Burundi exchange point chart.



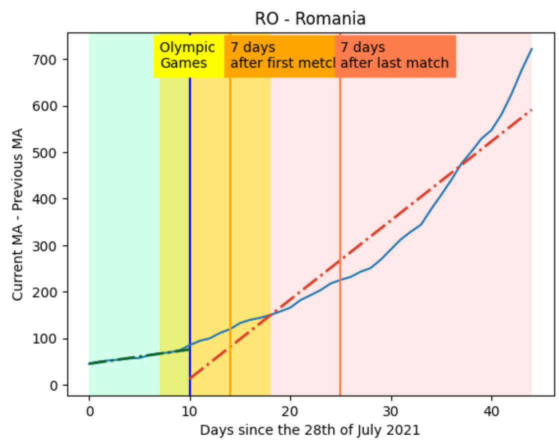
(a) Iceland exchange point chart.



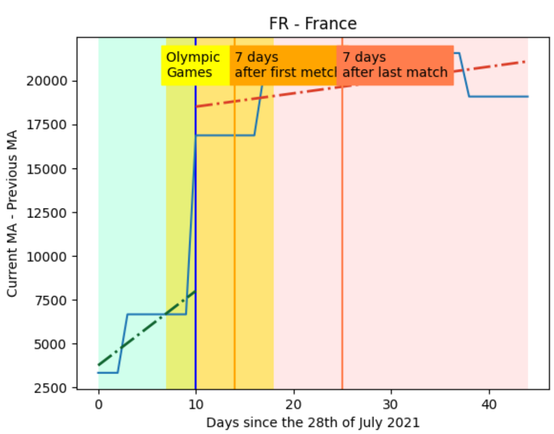
(b) Italy exchange point chart



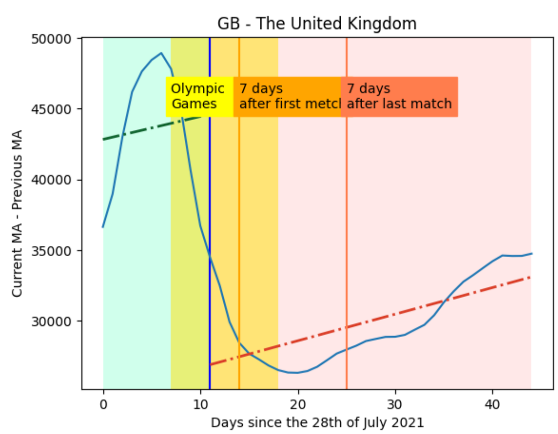
(a) Japan exchange point chart.



(b) Romania exchange point chart.



(a) France exchange point chart.



(b) The United Kingdom exchange point chart.

3.2 Autoregressive Integrated Moving Average(ARIMA) Prediction Models[2]

Previous analysis employed an algorithm based on Bayesian theory to discern transition moments in the trajectories of new COVID-19 infections. This methodology focussed its attention on the prerequisite and subsequent trend at the famous Olympic Games. However, we wanted to show an opportunity for further exploration of the data, taking a different approach and scrolling along a divergent perspective. In this scenario, the acuity moves from the detection of the reversal points to a division of the time history into four distinct traits: the pre-Olympic time span, the first week after the first race, the second week after the close, and two weeks after the close. This dual analytical strategy makes it possible to approach the analysis of a complex phenomenon from different angles, revealing unexplored facets in the data available. This subdivision is not accidental, but is driven by well-founded considerations. Previous studies have shown that the incubation period of COVID-19 is about one week[10]. As a result, if many people had contracted the virus during Olympic-related social gatherings, a significant increase in cases would be expected, highlighting a spike in reports of new cases between one and two weeks after the event.

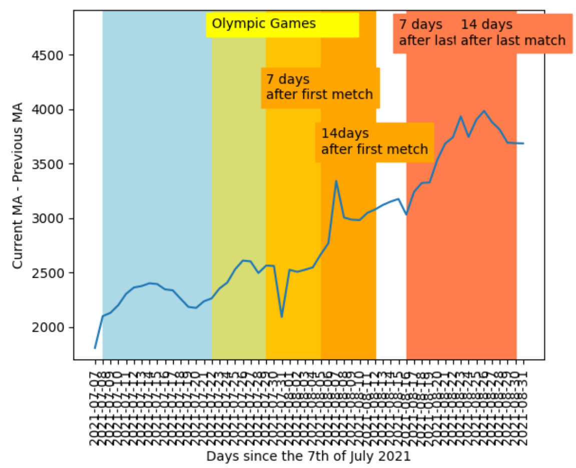


Figure 8: Chart with the areas of interest for the study.

The origin of this analysis is imbued with a seemingly simple question: what influence does the incubation period of an infectious disease exert on the moment when the climax in time sequences occurs? In addition, the crucial question arises as to whether the evolution and immunisation efforts in that given time frame mitigated the implications that relevant social events, such as the Olympics, could trigger. The results of these surveys show that, in fact, an increase in reports of new cases was found in the first week following the start of the first competitions, involving 64.70% of the states (with an average forecast error of less than 50% before the sporting event), which showed an increasing trend that differs by more than 50% from the forecast of our model. Similarly, after the closing of the races, in the first week, of those states, an increase of 48.76% of the states was observed, indicating an increasing trend. Otherwise, 11% of the states showed dynamics very similar to the prediction indicating that there was no event that led to a case upheaval of more than 10% of average error.

	10%	20%	50%
N.Paesi	40	61	121

Table 2: Table of countries that before the Olympic Games had average errors of the forecasts of minor infections of 10%, 20% and 50% of the 187 countries had vaccinations.

We show the following summary tables showing the trend of the average error as a percentage and the number of countries that have or have not fallen within that error range.

	Errore_medio_limit	7 day first metch	14 day first metch	7 day second metch	14 day second metch
0	10%	13	10	9	4
1	20%	19	17	15	11
2	50%	33	27	24	24

Table 3: Table of countries that before the Olympics had an average forecast error of less than 10% and their subsequent performance.

	Errore_medio_limit	7 day first metch	14 day first metch	7 day second metch	14 day second metch
0	10%	17	14	12	6
1	20%	26	22	20	14
2	50%	52	42	38	39

Table 4: Table of countries that before the Olympics had an average forecast error of less than 20% and their subsequent performance.

	Errore_medio_limit	7 day first metch	14 day first metch	7 day second metch	14 day second metch
0	10%	20	18	14	10
1	20%	31	29	24	20
2	50%	75	66	59	58

Table 5: Table of countries that before the Olympics had an average forecast error of less than 50% and their subsequent performance.

For example, we can observe that of the 40 states that before COVID had an average error of less than 10%, 36 after the second week after the end of the Olympic Games had forecasts with major errors and therefore that the model was unable to predict the increase in infections due to extraordinary events such as the possible correlation with the Olympic Games. Analysing the opposite extreme, of the 121 countries that before the first race had a model with the average forecast error of less than 50% only 58(47.93%) maintained the same error, while 56.06% showed a trend outside the model's forecast. Let's see below example graphs of the predictions prior to the event and the next trend with the detachment of the forecasts from the real cases. In ble is depicted the trend of real cases and in orange the prediction made by the model of that country taking into account the trend of vaccinations of the period before the races. The first one we observe is an example of the ideal case where we see a detachment of the forecasts right from the first week after the races and we see that the graph belongs to Japan, the country where all the Olympic Games took place.

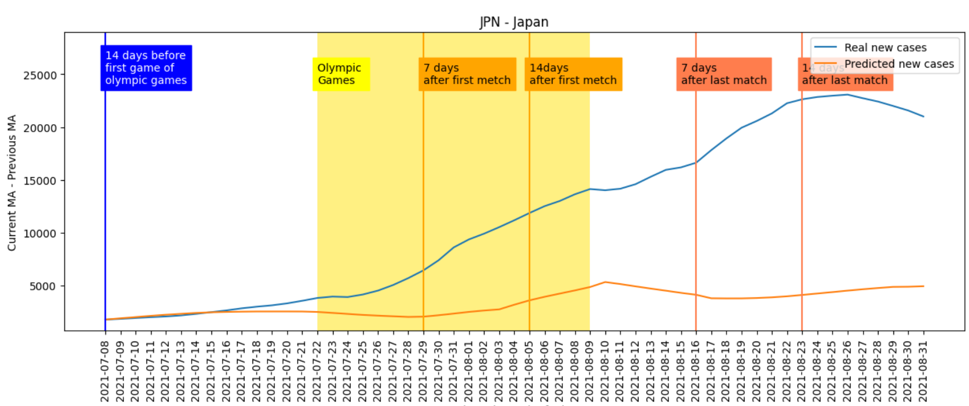


Figure 9: Japan prediction and real new daily cases.

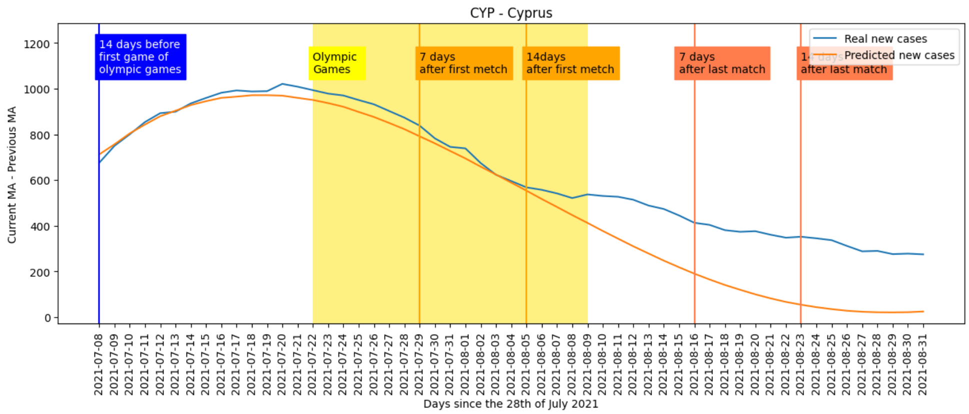


Figure 10: Cypro prediction and real new daily cases.

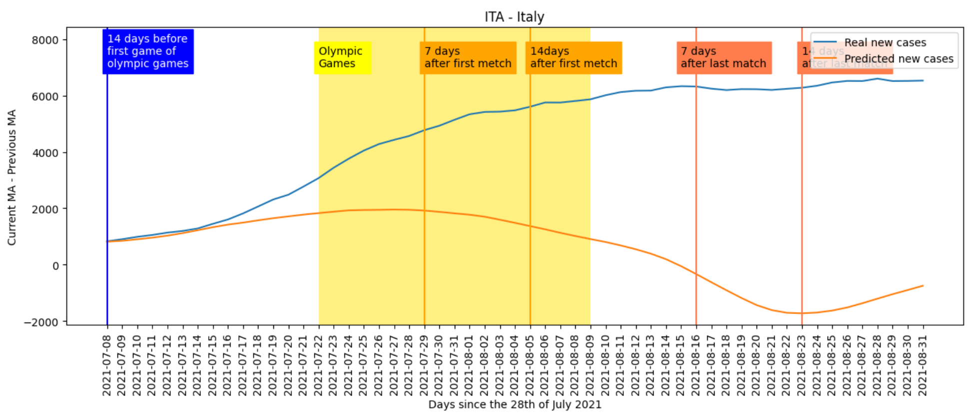


Figure 11: Italy prediction and real new daily cases.

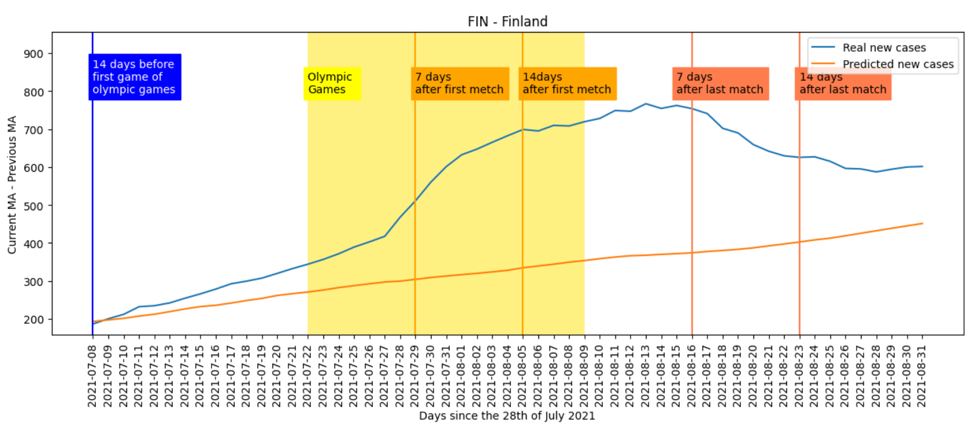


Figure 12: Finland prediction and real new daily cases.

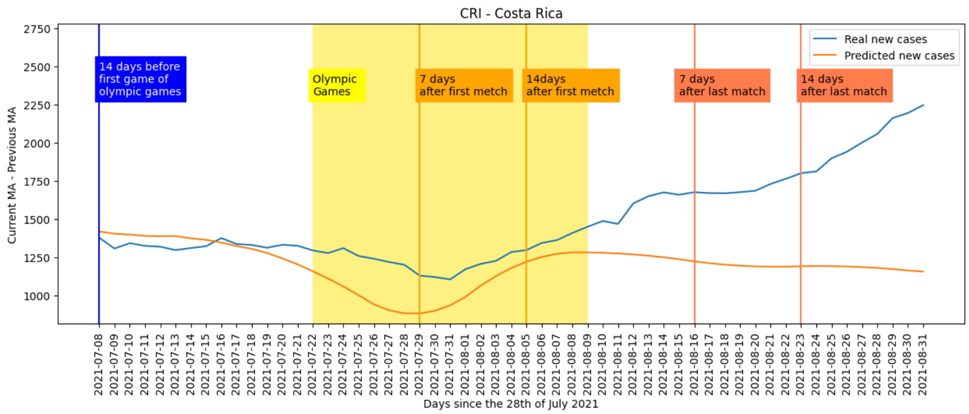


Figure 13: Costa Rica prediction and real new daily cases.

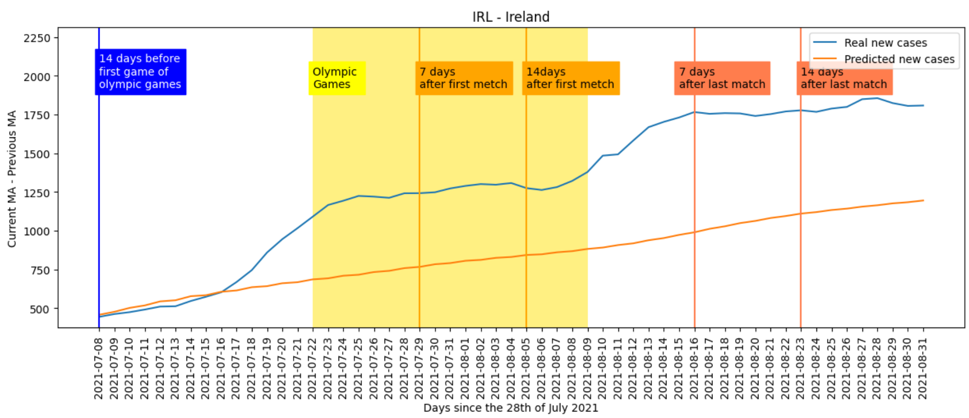


Figure 14: Irland prediction and real new daily cases.

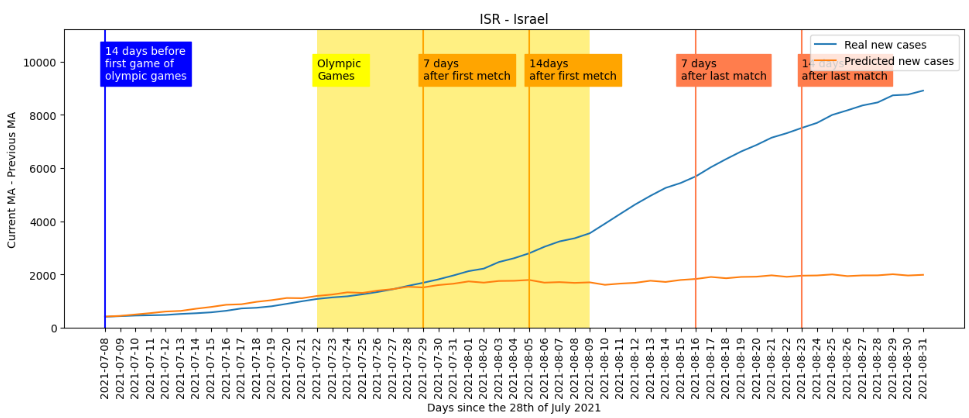


Figure 15: Israel prediction and real new daily cases.

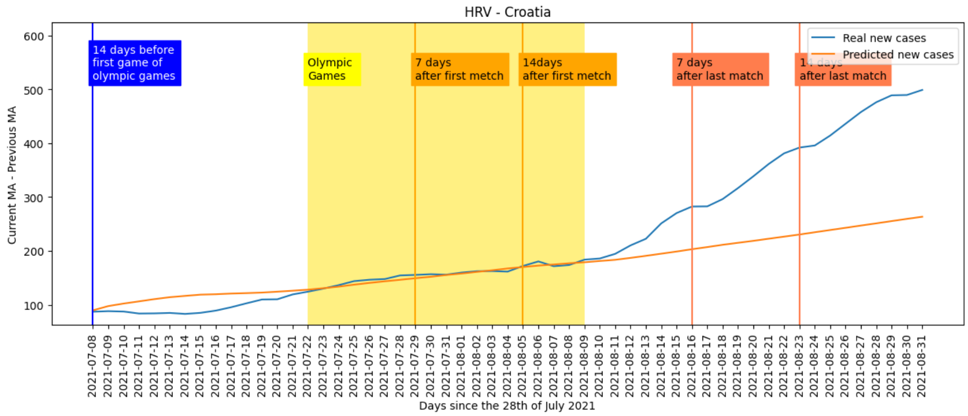


Figure 16: Croatia prediction and real new daily cases.

4 Discussion and Conclusion

This study constitutes a bivalent analysis of the impact of the global sporting event such as the Olympic Games on the trend of new COVID-19 cases at the state level.

A preliminary form of analysis employs the use of complex Bayesian methods in order to trace the instant in which a substantial deviation occurs in a time series. The thorough dissection of this process showed that, considering a general average, the time interval in which this deviation occurs for all 206 states and territories subject to the study is approximately 20.89 days from the start of the investigation or 13.89 days from the first competition. This reveal is inherently intriguing, as it clearly indicates that the Olympic event has generated a certain degree of effect on the analysed data. A significant aspect that emerged from the research is related to the fact that of the 206 states and territories examined, as many as 129 experienced a deterioration in conditions after the exchange point. This evolution was assessed by analysing the slope of a first degree polynomial, which adapts to the distribution of both pre- and post-exchange point data. Despite the period in which the competitions were held coincides with a time when nations and states were already fully aware of the accelerated spread of COVID-19 and had promptly implemented prevention measures even before the start of data collection, this was not enough to curb the emergence of positive dynamics much more relevant than the period before the start of the competitions.

The second analytical approach was performed through the application of statistical analysis models that uses time series data to better understand the dataset or to predict future trends, trained for each country on data of vaccination-free performed in that given state. Subsequently these models were used to see the trends of prefixed ranges. These intervals include the period before the first race of the Tokyo 2020 Olympics, one week and two weeks after and finally the week after the last race and after 2 weeks. The idea behind this analysis is that, given that COVID-19 incubation takes approximately 7-14 days, the main peak of cases should occur from the first week after the event, showing sharp deviations from the model's predictions. Therefore, the search for points of change can be avoided, focussing instead on the behaviour of average errors from the forecasts. For our analysis, accuracy ranges of the ARIMA models were made, it was calculated how many models before the Olympic Games had an average error from the forecast lower than three bands: 50%, 20%,10%. Of these we then saw how many countries showed behaviour not foreseen by our models. Let's say for example the 40 countries with models that before the Olympics had an average error of less than 10%, 33 of these countries in the second week after the last race showed an increase in cases with an average error from the forecasts of more than 50%.

Note, however, that these findings cannot be considered conclusive, as complex events such as a pandemic cannot be fully understood through these studies. As a result, some states may have data patterns that do not fully support such claims.

In summary, this research highlights the potential connection between significant sports events (along with related smaller activities) and a rise in the occurrence of contagious illnesses. This correlation becomes more evident in the week immediately following the occasion, during which over half of the analyzed regions displayed an uptick in the spread of these diseases.

References

[1] Olympic games tokyo 2020 watched by more than 3 billion people. https://olympics.com/ioc/news/olympic-games-tokyo-2020-watched-by-more-than-3-billion-people.

[2] Joshua Samuel Raj. R, S. Anantha Babu, Joy Winston James, and Ramchand Vedaiyan. Arima based time series analysis: Forecast covid-19 most vaccinated process and active cases classify using probability distribution curve rates (arimapdc). In *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, pages 546–551, 2021.

[3] Luca Casini and Marco Roccetti. A bayesian analysis of the inversion of the sars-cov-2 case rate in the countries of the 2020 european football championship. *Future Internet*, 13(8):212, 2021.

[4] Green pass eu. https://www.dgc.gov.it/web/.

[5] Covid-19 cases dataset owid.org. https://ourworldindata.org/coronavirus.

[6] World health organization - covid-19 - data. https://covid19.who.int/data.

[7] Olympic game calendar. https://www.barcalcio.net/olimpiadi-2020-programma-delle-gare-novita-discipline/#Olimpiadi_2021_calendario_delle_gare.

[8] Covid-19 vaccinations dataset kaggle.com. https://kaggle.com/datasets/gpreda/covid-world-vaccination-progress.

[9] Iso code 3166. https://www.iso.org/iso-3166-country-codes.html.

[10] J.A. Quesada, A. López-Pineda, V.F. Gil-Guillén, J.M. Arriero-Marín, F. Gutiérrez, and C. Carratala-Munuera. Incubation period of covid-19: A systematic review and meta-analysis. *Revista Clínica Española (English Edition)*, 221(2):109–117, 2021.

Appendix A

This appendix contains the link to a Google Drive folder containing all graphs for both the analysis techniques used in this study.
Link - https://drive.google.com/drive/folders/1YrPNt8FT5P7i-V64sv21dcxD5yEM6429?usp=share_link
GitHub code - https://github.com/DavideTalevi98/HDS_Olympic_Games_COVID