

به نام او

گزارش تمرین تحویلی فوت اضافه

شاهد رضوی زاده - ۹۹۱۰۴۶۲۷

نخست دیتاست مربوط به آمار فوتی‌های استان‌ها بر اساس ماه را وارد کردم و سپس دیتاست جدیدی ساختم که در آن برای هر استان و هر ماه-سال، میزان فوتی هایش مشخص شده بود. سپس دی ماه ۱۳۹۸ را به عنوان مرز بین دوران پیشاکرونا و پساکرونا در نظر گرفتم، و با توجه به این موضوع، برای هر ترکیب ماه-سال و استان، بر اساس ۵ داده اخیر یک مدل رگرسیونی ساده تعداد فوتی برحسب ماه-سال ساختم.

سپس با این مدل، مقدار پیش بینی شده برای سال‌های آتی یعنی ۱۳۹۹ و ۱۴۰۰ و ۱۴۰۱ را به دست آوردم و با مقدار فوتی‌های واقعی این مقادیر را مقایسه کردم. در صورتی که مقادیر واقعی به میزان خوبی outlier بودند (فاصله شان از مقدار پیش بینی شده حداقل به اندازه دوبرابر انحراف معیار باشد) آن گاه این اختلاف بین مقدار واقعی و پیش بینی شده را به عنوان فوت اضافی ناشی از کرونا در نظر گرفتم و به دیتاست جدیدی به نام result اضافه کردم. همچنین بدون در نظر گرفتن این شرط آخر، دیتاست result_2 رو نیز برای مراحل بعدی ساختم.

همچنین اگر مقدار p-value برای هر مدل رگرسیونی ساده‌ای که ساختم بالاتر از ۰٫۱ بود، در این صورت از مدل رگرسیونی صرف نظر می‌کردم و مقدار میانگین ۵ داده گذشته را به عنوان مقدار پیش بینی شده برای ماه-سال‌های آتی در دوران پساکرونا در نظر گرفتم. دلیل مقدار ۰٫۱ هم این بود که هم تعداد مشاهدات کم بود و از طرفی هم چون به هر حال حساسیت پیدا کردن مقدار فوت اضافی ناشی از کرونا بالاست، خواستم تا مرز p-value بیشتر از ۰٫۱ برای معناداری نتایج ما نشود.

برای ساخت نمودار حرارتی (که در فایل zip قرار گرفته است) نخست داده‌های فوت اضافی را نرمالایز کردم، به این صورت که نسبت فوت اضافی به تعداد فوتی‌ها در هر ترکیب ماه-سال و استان را حساب کردم و در متغیر ratio ذخیره کردم و براساس ratio ها مقایسه و در نهایت نمودار را ساختم. که همان طور که مشاهده می‌کنید، موج‌های بزرگ کرونایی که در کل کشور ایجاد شد، در شکل هم قابل مشاهده هستند.

سپس با استفاده از داده‌های result، مقادیر فوت اضافه هر استان و در نهایت کل کشور را به دست آوردم. جدول مربوط به سوال ۲ یعنی فوتی‌های اضافه هر استان در فایل با نام result_prov در zip قرار گرفته است. همچنین تخمین من نشان داد که به خاطر کرونا، در کل کشور ۲۹۳۲۰۰ نفر فوت کرده اند.

در آخر نیز برای سوال آخر، عملکرد ratio بین شهریور ماه ۱۴۰۰ و ۱۴۰۱ را برای هر استان مقایسه کردم و اختلاف این‌ها را به عنوان پیشرفت هر استان در نظر گرفتم. از آن جایی که برای استان سیستان و بلوچستان و چهارمحال و بختیاری، نقص داده برای این مقایسه وجود داشت، به ناچار از داده‌های result_2 استثنائاً کمک گرفتم و به همین جهت result_2 را ساختم و از این جهت نتیجه عملکرد این دو استان به معناداری نتایج دیگر استان‌ها نیست.