**Exploratory Part**

**Preprocessing:**

X dataset is of the house prices in Connecticut each month from 2020 to 2022. We have generated a new column which indicates the percentage change in the House price per month.

In both the Cases and Vaccination datasets, mean of daily count per month has been used as monthly count data.

All the data available has been normalized using min-max normalization.

**a. Pearson Correlation for case, vaccine and house rate data between Jan 2021 to May 2021 (Connecticut)**

Below are a few events listed which have impacted the data in consideration

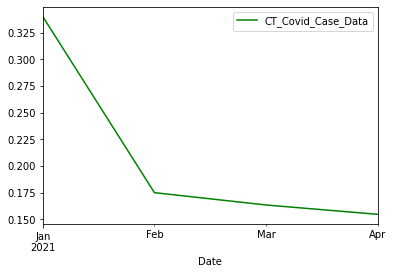
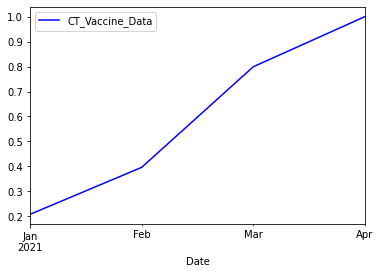
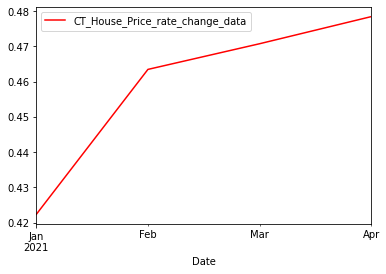
**January 4**: Healthcare workers began receiving scheduled second vaccination doses for those who had already received one dose.

**February 1**: The mandatory closing time for dine-in restaurants was moved from 10 PM to 11 PM. Mask guidelines, social distancing recommendations, and the 50% capacity limit on houses of worship were left unchanged, but the numerical cap on group size was lifted.

**February 8**: As of this date, more than 50% of the population 75 years old or older was partially vaccinated, leading to the announcement that vaccinations would be opened to those 65 or older.

**May 3**: Connecticut was the first state to fully vaccinate 50% of its adult population.

**Graph for each dataset:**



Calculation of Pearson Correlation between 3 possibilities:

**Case 1: Mean Monthly Covid Case Count vs Mean Monthly Vaccination Count**

H0: Mean Monthly Covid Case Count and Mean Monthly Vaccination Count are not linearly correlated

H1: Mean Monthly Covid Case Count and Mean Monthly Vaccination Count are linearly correlated

Calculated Value = -0.78351845

**As |-0.78351845| > 0.5, We reject the null hypothesis**

Mean Monthly House Price Change Rate and Mean Monthly Covid Case Count have a **negative linear correlation**

**Case 2: Mean Monthly House Price Change Rate vs Mean Monthly Covid Case Count**

H0: Mean Monthly House Price Change Rate and Mean Monthly Covid Case Count are not linearly correlated

H1: Mean Monthly House Price Change Rate and Mean Monthly Covid Case Count are linearly correlated

Calculated Value = -0.98832128

**As |**-**0.98832128| > 0.5, We reject the null hypothesis**

Mean Monthly House Price Change Rate and Mean Monthly Covid Case Count have a **negative linear correlation**

**Case 3: Mean Monthly House Price Change Rate vs Mean Monthly Vaccination Count**

H0: Mean Monthly House Price Change Rate and Mean Monthly Vaccination Count

are not linearly correlated

H1: Mean Monthly House Price Change Rate and Mean Monthly Vaccination Count

are linearly correlated

Calculated Value = 0.86542085

**As |0.86542085| > 0.5, We reject the null hypothesis**

Mean Monthly House Price Change Rate and Mean Monthly Vaccination Count have a **positive linear correlation**

**Results**:

House price rate change and Vaccination count have a positive linear correlation with each other and a negative linear correlation with Covid Case Count between January 2021 and May 2021.

According to the timeline of COVID in Connecticut during this time range (mentioned earlier), it can be deduced that increase in vaccination has positively impacted the Connecticut real estate industry and has led to reduction in the Covid cases.

All 3 data sets are linearly correlated to one another.

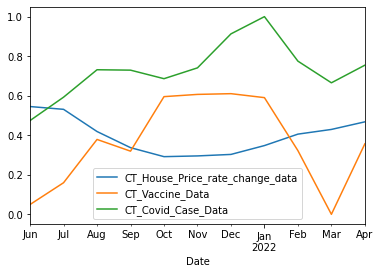
**b. Pearson Correlation for case, vaccine and house rate data between June 2021 to May 2022 (Connecticut)**

Below are a few events listed which have impacted the data in consideration

##### **June 1, 2021**: The Delta variant, first identified in India in late 2020, becomes the dominant variant in the U.S. The variant kicks off a third wave of infections during the summer of 2021. The present vaccination was about 90% effective against the delta variant and the cases kept increasing

##### **November 26, 2021**: World Health Organization classifies a new variant, Omicron, as a variant of concern after it was first reported by scientists in South Africa. The variant has several mutations in the spike protein that concern scientists around the world. The present vaccination was about not effective against the omicron variant and hence we see the spike in cases

**Graph for each dataset:**

****

Calculation of Pearson Correlation between 2 possibilities:

**Case 1: Mean Monthly House Price Change Rate vs Mean Monthly Covid Case Count**

H0: Mean Monthly House Price Change Rate and Mean Monthly Covid Case Count are not linearly correlated

H1: Mean Monthly House Price Change Rate and Mean Monthly Covid Case Count are linearly correlated

Calculated Value = ​​-0.647678754

**As |-0.647678754| > 0.5, We reject the null hypothesis**

Mean Monthly House Price Change Rate and Mean Monthly Covid Case Count have a **negative linear correlation**

**Case 2: Mean Monthly House Price Rate Change vs Mean Monthly Vaccination Count**

H0: Mean Monthly House Price Change Rate and Mean Monthly Vaccination Count

are not linearly correlated

H1: Mean Monthly House Price Change Rate and Mean Monthly Vaccination Count

are linearly correlated

Calculated Value = -0.8194726285

**As |-0.8194726285| > 0.5, We reject the null hypothesis**

Mean Monthly House Price Change Rate and Mean Monthly Vaccination Count have a **negative linear correlation**

**Results**:

House prices and vaccination counts have a negative linear correlation. This may be attributed to the advent of omicron and delta variants and hence this trend is different to the one seen in part a.

**c. Chi-Square testing to check independence of data samples**

1. **Check the independence between housing price change rate and administration of different vaccines in the state of Connecticut.**

**H0**: Housing price change rate (HPCR) is independent of vaccine administration in the state of Connecticut

**H1**: Housing price change rate (HPCR) is dependent on vaccine administration in the state of Connecticut

|  | Pfizer | Janssen | Moderna | Total |
| --- | --- | --- | --- | --- |
| Max house price rate change | 247857 | 16575 | 102437 | 366869 |
| Min house price rate change | 270293 | 6398 | 85013 | 361704 |
| Total | 518150 | 22973 | 187450 | 728573 |

**Qscore** = 7063.22

**Degree of Freedom** = (2-1) \* (3-1)

**p-value** = 0.0 (nan) < 0.05

Hence , we reject the null hypothesis

HPCR is not independent of the vaccines administered in Connecticut.

1. **Check the independence between housing price change rate and deaths due to covid in the states of California, Texas, Florida and New York.**

California, Texas, Florida and New York have been the 4 states with the highest deaths during covid pandemic. The months selected for chi square are the ones which have the maximum and minimum

HPCR.

**H0**: Housing price change rate (HPCR) is independent of deaths due to coronavirus.

**H1**: Housing price change rate (HPCR) is dependent on deaths due to coronavirus.

|  | California Deaths | Texas Deaths | Florida Deaths | New York Deaths | Total |
| --- | --- | --- | --- | --- | --- |
| Max HPCR month | 1013 | 535 | 903 | 151 | 2602 |
| Min HPCR month | 2850 | 6006 | 2665 | 793 | 12314 |
| Total | 3863 | 6541 | 3568 | 944 | 14916 |

**Qscore** = 751.21

**Degree of Freedom** = (2-1) \* (4-1)

**p-value** = 0.0 < 0.05

Hence , we reject the null hypothesis

HPCR is not independent deaths due to coronavirus based on data from top death experiencing cities and months with minimum and maximum HPCR

**d. Pearson Correlation testing between deaths during covid and Housing Price Change Rate**

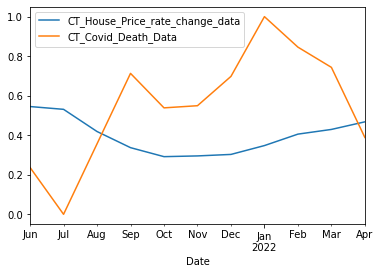
Below are a few events listed which have impacted the data in consideration

**June 1, 2021**: The Delta variant, first identified in India in late 2020, becomes the dominant variant in the U.S. The variant kicks off a third wave of infections during the summer of 2021.

**November 26, 2021**: World Health Organization classifies a new variant, Omicron, as a variant of concern after it was first reported by scientists in South Africa. The variant has several mutations in the spike protein that concern scientists around the world. [[LINK](https://www.census.gov/content/dam/Census/library/stories/2022/03/united-states-deaths-spiked-as-covid-19-continued-figure-3.jpg)]

Deaths began to drop In January 2021. By April that year, the United States saw a decline in deaths corresponding to increases in vaccinations. And that June, death levels returned to near June 2019 levels. This decline was short-lived, though, as the Delta variant caused another spike in deaths in July and August. Deaths dropped slightly in the fall, but then the Omicron variant emerged just as holiday travel picked up in December 2021. [[LINK](https://www.census.gov/library/stories/2022/03/united-states-deaths-spiked-as-covid-19-continued.html)]

**Graph for each dataset:**

****

**Case: HPCR vs Death Count**

H0: House Price Change Rate and Covid Death Count are not linearly correlated

H1: House Price Change Rate and Covid Death Count are linearly correlated

Calculated Value = ​​0.65040010

**As |-0.65040010| > 0.5, We reject the null hypothesis**

Mean Monthly House Price Change Rate and Mean Monthly Covid Case Count have a **negative linear correlation**

**Results**:

House prices and death counts have a negative linear correlation. This may be attributed to the advent of omicron and delta variants. Its potential justification is mentioned in the event description.