Kantsir Solomia, Muravetska Yaryna

D.A.W.N. vs Coronavirus

The theme of the project

The theme of our project is to examine the relationship between Covid-19 and companies from 4 different branches. (Domino's Pizza - now it is a food delivery, Activision - online games, Walmart - worldwide supermarket, Netflix - online movie streaming)
We have the hypothesis that Covid-19 has positively influenced those online services (games and streaming), delivery services, and supermarkets. During our research we want to find out if this hypothesis can be accepted.

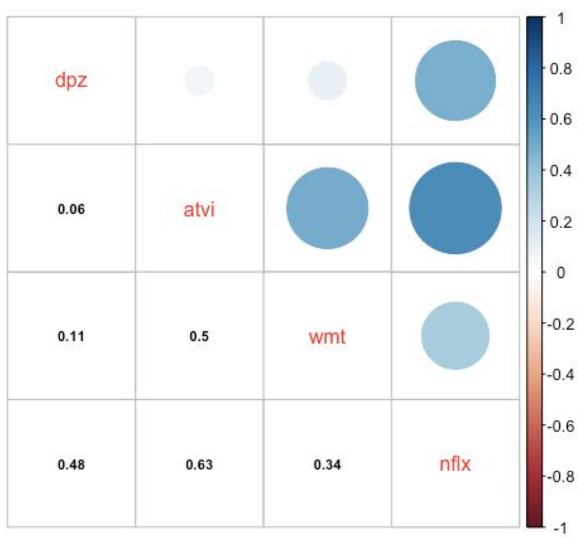
Aim of the project

Taking into account that online platforms, supermarkets, and delivery services in nowadays situation are getting more and more on-demand, we want to understand if there is a significant impact of COVID-19 on stocks of D.A.W.N. In our case D.A.W.N. companies are representatives of most demanded areas. We will have a model for each company and will figure out which variables (other companies' stocks and COVID19 cases) will be significant in those models and their coefficients (effect on the dependent variable). Also we want to predict stock prices of those companies in the nearest future, based on our model and historical data.

Research process

Step 1

Firstly, we decided to examine the correlation between all the companies' stocks. dpz - stands for Domino's Pizza; atvi - Activision; wmt - Walmart; nflx - Netflix



According to the plot we see that positive correlation is between Domino's and Netflix, Activision and Netflix, Activision and Walmart. That means that we can assume some kind of positive relationship between stocks of those companies.

In collected data on Covid-19 we had to decide which variable we will check for relationship with stocks (we had total cases of illness, only deaths, cases per day etc). We selected two most relevant (cases, deaths) and made a correlation matrix from which we conducted that although both have strong correlation, the deaths should be preferred in models with atvi, wmt, nflx because of better correlation coefficient. In the regression with dpz, we will use variable 'cases', because of stronger correlation.

	dpz	atvi	wmt	nflx
cases	0.4928090	0.08063486	0.5015717	0.4335133
deaths	0.4636085	0.12313104	0.5418441	0.4343442

Step 2

We will make 8 different OLS regressions (2 for each company) to examine if there is a significant relationship between stocks of a particular company and other factors.
First model will include only 1 independent variable - deaths/cases and the second one will include deaths/cases + the stocks of other companies. These 2 regressions will help us to find a more precise/explaining model and then make a conclusion based on the chosen one.

Netflix

1. We want to examine how Netflix stocks are dependent on deaths from Covid-19

   $H_0 =$ *there is no significant relationship between Netflix stocks and deaths*
   $H_1 =$ *there is significant relationship between Netflix stocks and deaths*

   $nflx = \beta_0 + \beta_1 deathes + \varepsilon$

   Having run an OLS regression we have the next coefficient estimation:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.515e+02  2.654e+00 132.480  < 2e-16 ***
deaths      3.616e-04  8.964e-05   4.034 0.000138 ***
---
```

   From here we see that deaths is statistically significant variable with $\beta_1 = 3.616e^{-04}$ and p-value = < 2e-16 , so there is a significant relationship between deaths and nflx. p value <0.05, so we reject H_0
   P-value of the model : 0.0001378 -> statistically significant. Reject H0

2. We want to examine how Netflix stocks are dependent on deaths from Covid-19, dpz, atvi, wmt
   $nflx = \beta_0 + \beta_1 deaths + \beta_2 dpz + \beta_3 atvi + \beta_4 wmt + \varepsilon$

   $H_0 =$ *there is no significant relationship between Netflix stocks and deaths, dpz, atvi, wmt*
   $H_1 =$ *at least one* $\beta i \neq 0,\ i = 1, ..., k$

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.739e+01  6.363e+01   1.059 0.293378
deaths       2.573e-04  8.497e-05   3.029 0.003487 **
dpz          2.355e-01  6.266e-02   3.758 0.000361 ***
atvi         5.661e+00  7.317e-01   7.736 7.19e-11 ***
wmt         -1.075e+00  5.800e-01  -1.854 0.068174 .
---
```

From summary we see that all variables are statistically significant. We reject H_0
'deaths' have positive dependence with Ntflx stocks -> 1 death cause 2.573e-04 units increase in stocks
dpz (Domino's stocks) and atvi (Activision) also have positive dependence with Nflx stocks with coefficients 2.355e-01 and 5.661e+00, respectively.
Walmart stocks have negative correlation with Ntflx stocks -> 1 unit change in Wlmrt stocks cause 1.075e+00 units decrease in Ntflx stocks
Model is statistically statistically significant with p-value: 2.36e-14. Reject H0.

Domino's Pizza

1. This model will be regressed to find the dependence of Domino's Pizza stocks on Covid-19 cases.

$$dpz = \beta_0 + \beta_1 cases + \varepsilon$$

$H_0 =$ *there is no significant relationship between Domino's stocks and cases*
$H_1 =$ *there is significant relationship between Domino's stocks and cases*

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3.038e+02	3.574e+00	85.003	< 2e-16	***
cases	3.278e-05	6.919e-06	4.738	1.09e-05	***

'cases' is statistically significant.1 increase in cases causes 3.278e-05 units change in stocks.
The model has p value =1.094e-05, that is statistically significant (p<= 0.05), but R-squared: 0.2429
, that is pretty low. That means that model doesn't explain much of variation of the data but it is significant. Reject H0.

2. In this model we want to examine the relationship between Domino's stocks , Covid-19 cases and other listed companies stocks.

$$dpz = \beta_0 + \beta_1 cases + \beta_2 nflx + \beta_3 atvi + \beta_4 wmt + \varepsilon$$

$H_0 =$ *there is no significant relationship between Domino's stocks and cases, nflx, atvi, wmt* ($\beta_1 = \beta_2 = \ldots = \beta_k = 0$)
$H_1 =$ *at least one* $\beta i \neq 0,\ i = 1, \ldots, k$

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3.276e+02	1.022e+02	3.205	0.002069	**
cases	2.348e-05	9.046e-06	2.595	0.011605	*
nflx	6.967e-01	1.995e-01	3.492	0.000855	***
atvi	-2.609e+00	1.776e+00	-1.469	0.146380	
wmt	-9.832e-01	1.022e+00	-0.962	0.339659	

We have only two statistically significant variables : 'cases' and 'nflx', that means that cases have positive correlation with Domino's stocks - 1 change in cases causes 2.348e-05 units of change in stocks, and with Netflix stocks -1 change in stocks causes 6.967e-01 units of change in Domino's stocks.
Reject H0

Walmart

1. Here we want to see if Walmart stocks have a relationship to Covid-19 deaths.

   $wmt = \beta_0 + \beta_1 deaths + \varepsilon$

   $H_0 =$ *there is no significant relationship between Walmart's stocks and deaths*

   $H_1 =$ *there is significant relationship between Walmart's stocks and deaths*

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.152e+02  4.437e-01 259.644  < 2e-16 ***
deaths      8.084e-05  1.499e-05   5.394 8.86e-07 ***
---
```

   So, Walmart stocks and Covid-19 deaths are positively correlated.
   Coefficient is statistically significant with p-value=8.86e-07. Reject H0

2. The model is more extensive and explaining

   $wmt = \beta_0 + \beta_1 deaths + \beta_2 atvi + \beta_3 nflx + \beta_4 dpz + \varepsilon$

   $H_0 =$ *there is no significant relationship between Walmart's stocks and deaths, atvi, nflx, dpz*

   $H_1 =$ *at least one* $\beta i \neq 0,\ i = 1, ..., k$

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.232e+01  8.520e+00   9.663 2.53e-14 ***
deaths       9.299e-05  1.474e-05   6.308 2.57e-08 ***
atvi         8.862e-01  1.762e-01   5.029 3.92e-06 ***
nflx        -4.537e-02  2.448e-02  -1.854   0.0682 .
dpz         -1.242e-02  1.408e-02  -0.882   0.3810
```

   Only ‘deaths’ and ‘atvi’ are statistically significant with p values 2.57e-08 and 3.92e-06 Respectively.
   This model explains a lot of variation within the data and is significant (R-squared: 0.5306 and p-value: 1.86e-10)
   Reject H0

Activision

1. Regression with this model will show if Activision stocks and Covis-19 deaths are dependent.

   $atvi = \beta_0 + \beta_1 deaths + \varepsilon$

   $H_0 =$ *there is no significant relationship between Activision's stocks and deaths*

   $H_1 =$ *there is significant relationship between Activision's stocks and deaths*

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.938e+01  3.453e-01 171.969   <2e-16 ***
deaths      1.211e-05  1.166e-05   1.038    0.303
```

   Deaths variable is statistically insignificant because of high p-value.
   What’s more model is pretty bad, because of extremely low R-squared and too high p-value

```
Multiple R-squared:  0.01516,	Adjusted R-squared:  0.001092
F-statistic: 1.078 on 1 and 70 DF,  p-value: 0.3028
```

Reject H1

2. This model will help to understand if there is any relationship between ACtivision stocks and Covid-19 deaths, nflx, wmt, dpz

$$atvi = \beta_0 + \beta_1 deaths + \beta_2 nflx + \beta_3 wmt + \beta_4 dpz + \varepsilon$$

$$H_0 = \textit{there is no significant relationship between Activision's stocks and deaths, nflx, wmt, dpz}$$
$$H_1 = \textit{at least one } \beta i \neq 0,\ i = 1,\ ...,\ k$$

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.235e+00  7.785e+00  -0.159 0.874464
deaths      -3.567e-05  1.009e-05  -3.534 0.000747 ***
nflx         8.335e-02  1.077e-02   7.736 7.19e-11 ***
wmt          3.092e-01  6.149e-02   5.029 3.92e-06 ***
dpz         -1.413e-02  8.187e-03  -1.726 0.088918 .
```

Having made an OLS regression we see that there are 3 statistically significant variables:
Deaths, hflx and wmp. Domino's stocks do not have any significant impact on Activision stocks.

```
Residual standard error: 1.655 on 67 degrees of freedom
Multiple R-squared:  0.6229,    Adjusted R-squared:  0.6004
F-statistic: 27.67 on 4 and 67 DF,  p-value: 1.411e-13
```

The model is pretty exhaustive:

model explains a lot of variation within the data and is significant

Reject H0

Step 3

On this step we will make predictions of Covid-19 deaths and cases and company's stocks with the help of the Autoregression Model. An AR model will predict future behavior based on past behavior. Some of the models will include other predictor variables.

In R code we did AR1 with 2 methods for better precision.

We will conduct 3 Ar1 regressions (each with different predictor variables) for each company.

Then we will choose the model that predicts the best (comparing predicted values with the actual ones)

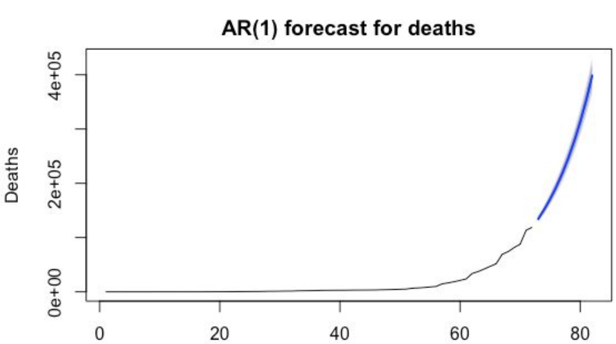
Death Rate AR1 model

- $deaths_{t} = \beta_0 + \beta_1\ deaths_{t-1} + \varepsilon$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 162.1279   359.8351   0.451    0.654
deaths_lags   1.1278     0.0137  82.307   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2707 on 69 degrees of freedom
Multiple R-squared:  0.9899,    Adjusted R-squared:  0.9898
F-statistic:  6774 on 1 and 69 DF,  p-value: < 2.2e-16
```

Summary of the model says that deaths_lags coefficient is statistically significant.
Also, the model is significant with small p-value (<0.05) , R-squared that is close to 1. But we can notice very big RSE.



Here we see the prediction plot of the given model. All the precise numbers can be found in the table in the R file. Plot shows an extreme increase in Covis-19 deaths during the next 10 days. To be

precise, point forecast predicts the rise from 134198.2 to 398464.8 deaths.
Also, we have the prediction interval for the Ar1 model. (Table with detailed values is in R file)
During the next 10 days (starting from 15.04.2020) 80% of deaths will lie within the interval (130778.6;419282.3) and 95% of deaths will lie within (128968.3; 430302.4)

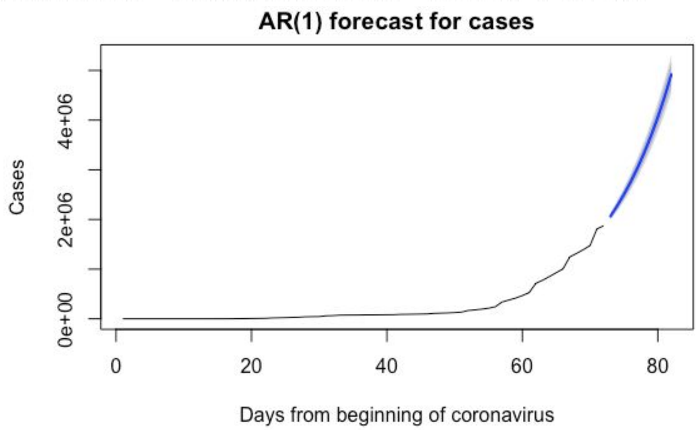
Cases AR1 model

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.575e+03  5.226e+03   0.493    0.624
cases_lags  1.100e+00  1.111e-02  99.036   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 38030 on 69 degrees of freedom
Multiple R-squared:  0.993, Adjusted R-squared:  0.9929
F-statistic:  9808 on 1 and 69 DF,  p-value: < 2.2e-16
```

- $cases_t = \beta_0 + \beta_1 cases_{lags} + \varepsilon$

From summary we see that the model is quite significant (small p-value and high R-squared).
$cases_{lags}$ variable is also statistically significant.



On the prediction plot one can see the rise in the number of cases. According to point forecast cases will increase from 2064003 to 4919649 in the next 10 days. During the next 10 days (starting from 15.04.2020) 80% of cases will lie within the interval (2015952;5171200) and 95% of cases will lie within (1990516; 5304362)

Netflix stocks AR1

- -With Ar1 model below we will predict **Netflix stocks** based only on previous historical data of its stocks. This model is quite simple, but we will use it in further regressions.
  $nflx_{level} = \beta_0 + \beta_1 nflx_{lags} + \varepsilon$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 50.63190   24.59183   2.059   0.0433 *
nflx_lags    0.86116    0.06904  12.473   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 12.27 on 69 degrees of freedom
Multiple R-squared:  0.6927,    Adjusted R-squared:  0.6883
F-statistic: 155.6 on 1 and 69 DF,  p-value: < 2.2e-16
```

Summary says that coefficient is statistically significant as well as the model itself due to p-values. Model doesn't explain much of the variation.

Here is a prediction graphic for Nflx stock prices for a given AR1 model.

We see that stocks mostly will fall in the next 10 days.
Dark grey area is 95% prediction interval, light grey - 80%.
Approximately, 80% of stocks will change from -45 to -17 units. And 95% will change on -50 -- -9 units.

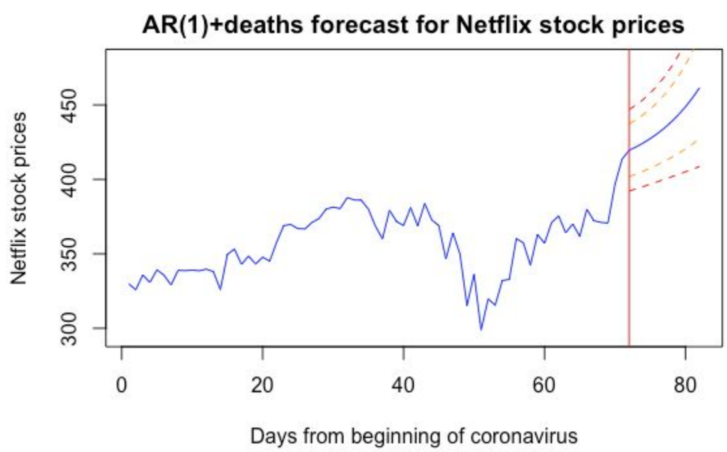
-More advanced **model for Nflx stocks** prediction is with one more predictor variable deaths_lags

$$nflx_{level} = \beta_0 + \beta_1 nflx_{lags} + \beta_2 deaths_{lags} + \varepsilon$$

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.932e+01  2.512e+01   2.760  0.00742 **
nflx_lags   8.036e-01  7.124e-02  11.280  < 2e-16 ***
deaths_lags 1.505e-04  6.407e-05   2.349  0.02174 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.88 on 68 degrees of freedom
Multiple R-squared:  0.7158,	Adjusted R-squared:  0.7074
F-statistic: 85.64 on 2 and 68 DF,  p-value: < 2.2e-16
```

This model is also statistically significant with the same p-value as the previous one, but with higher R-squared that means that the model describes a lot of variation.



Here is a prediction graphic for Nflx stock prices(after the red line).

Detailed precise numbers can be seen in the table in R file.
Orange dotted line stands for 95% prediction interval, red - 80%.
Approximately, 80% of stocks will rise on 20-40 units, 95% - on 15-50 units in the next 10 days. (from 14.04.2020)

-The third **Ar1 model for nflx** includes also other companies' previous (lagged) stocks

$$nflx\_level = \beta_0 + \beta_1\ nflx\_lags + \beta_2 deaths\_lags + \beta_3 atvi\_lags + \beta_4 wmt\_lags + \beta_5 dpz\_lags + \varepsilon$$

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.845e+02  5.480e+01   3.367  0.00128 **
nflx_lags    9.518e-01  1.037e-01   9.183 2.37e-13 ***
deaths_lags  2.186e-04  7.901e-05   2.767  0.00736 **
atvi_lags   -9.512e-01  8.488e-01  -1.121  0.26654
wmt_lags    -6.528e-01  5.058e-01  -1.291  0.20138
dpz_lags    -1.167e-01  5.821e-02  -2.004  0.04921 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 11.48 on 65 degrees of freedom
Multiple R-squared:  0.7464,	Adjusted R-squared:  0.7268
F-statistic: 38.25 on 5 and 65 DF,  p-value: < 2.2e-16
```

Statistically significant coefficients are $nflx\_lags$, $deaths\_lags$, dpz_lags
The model is statistically significant and explains a lot of variation. Has a bit high RSE.

Here is a prediction graphic for Nflx stock prices(after the red line). Detailed precise numbers can be seen in the table in R file.

We see that Nflx stock prices tend to grow extremely.Approximately, stocks will grow on 40 units in the next 10 days.

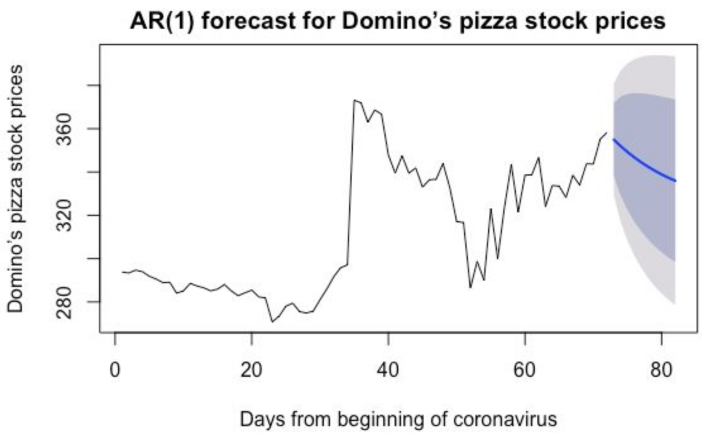
Domino's pizza Ar1 model

- $dmz_{level} = dmz_{lags} + \varepsilon$

Model is statistically significant as well as the dpz_lags variable.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.01783   16.91383   1.716   0.0907 .
dpz_lags     0.90981    0.05403  16.840   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Residual standard error: 13.39 on 69 degrees of freedom
Multiple R-squared:  0.8043,    Adjusted R-squared:  0.8015
F-statistic: 283.6 on 1 and 69 DF,  p-value: < 2.2e-16
```



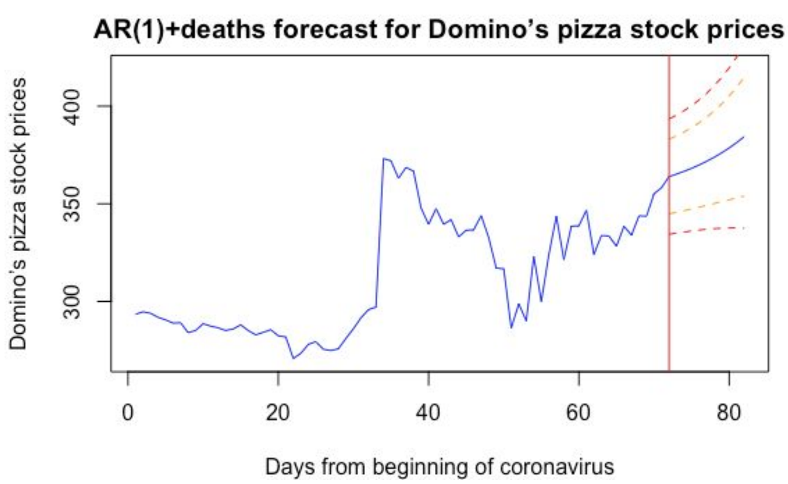
According to the plot, we have a pretty wide range of stocks movement, so it is not very precise. In general, stocks will fall on 20 units (point forecast), but a rise also can be expected. 80% of stocks can change from -40 to 2 units. 95% of stocks can change on -50 to 13 units

-The second AR1 model for Domino's pizza stocks has 2 predictor variables dmz_{lags} and $cases\_lags$, so we assume that it is more precise. $dmz_{level} = \beta_0 + \beta_1 dmz_{lags} + \beta_2\, cases_{lags} + \varepsilon$

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.091e+01  1.846e+01   2.216     0.03 *
dpz_lags    8.666e-01  6.048e-02  14.327   <2e-16 ***
cases_lags  6.713e-06  4.378e-06   1.533     0.13
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.26 on 68 degrees of freedom
Multiple R-squared:  0.8108,    Adjusted R-squared:  0.8053
F-statistic: 145.7 on 2 and 68 DF,  p-value: < 2.2e-16
```

Actually, it is almost the same as the previous one due to p-value and R squared. And here $cases_{lags}$ is not statistically significant.



According to the point forecast dpz stocks will have a rise of 20 units. 80% of stocks lie within the prediction interval (344; 409) in the next 10 days.
95% of stocks lie within the prediction interval (334; 425) in the next 10 days.

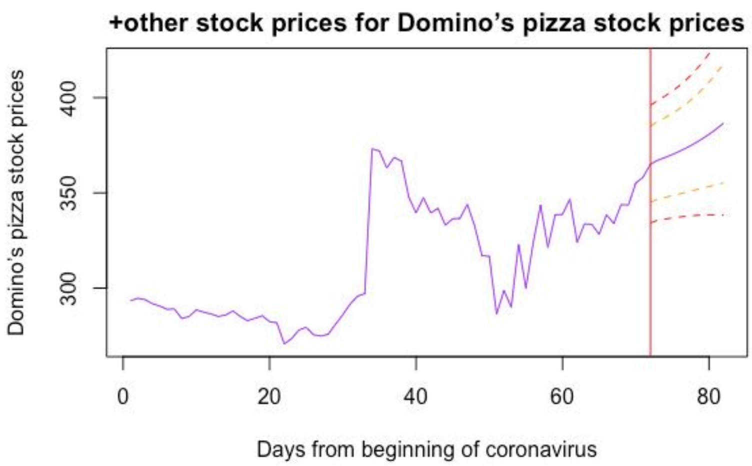
-The third Ar1 model includes stocks' of other companies as predictor variables.

$$dmz_{level} = \beta_0 + \beta_1 dmz_{lags} + \beta_2\ cases_{lags} + \beta_3 atvi\_lags + \beta_4 wmt\_lags + \beta_5 nflx\_lags + \varepsilon$$

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.243e+01  6.281e+01   0.835    0.407
dpz_lags     8.292e-01  6.837e-02  12.128   <2e-16 ***
cases_lags   6.817e-06  5.408e-06   1.260    0.212
atvi_lags   -1.767e-01  1.010e+00  -0.175    0.862
wmt_lags    -2.723e-01  5.850e-01  -0.465    0.643
nflx_lags    1.188e-01  1.223e-01   0.971    0.335
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.4 on 65 degrees of freedom
Multiple R-squared:  0.8154,	Adjusted R-squared:  0.8012
F-statistic: 57.43 on 5 and 65 DF,  p-value: < 2.2e-16
```

According to the summary of regression, we have the only significant variable - dpz_lags. The model itself is statistically significant (p-value <0.05), explains a lot of variation (R squared = 0.8) but has a bit high RSE.



From the plot we see that stocks mostly tend to grow.
Due to point forecast Domino's stocks will grow on 20 units.
80% of stocks lie within the prediction interval (345; 417) in the next 10 days.
95% of stocks lie within the prediction interval (334; 434) in the next 10 days.

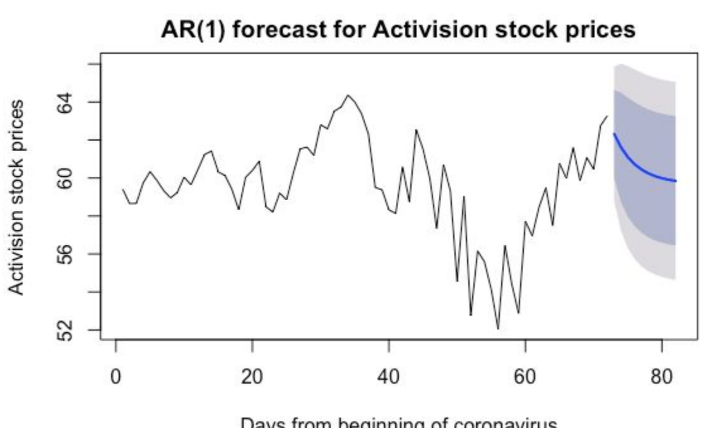
Activision stock's model

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.92431    5.03001   3.166   0.0023 **
atvi_lags    0.73321    0.08448   8.679 1.15e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Residual standard error: 1.836 on 69 degrees of freedom
Multiple R-squared:  0.5219,	Adjusted R-squared:  0.515
F-statistic: 75.32 on 1 and 69 DF,  p-value: 1.148e-12
```

$$atvi_{level} = \beta_0 + \beta_1 atvi_{lags} + \varepsilon$$

Model doesn't explain much of variation, but is statistically significant.



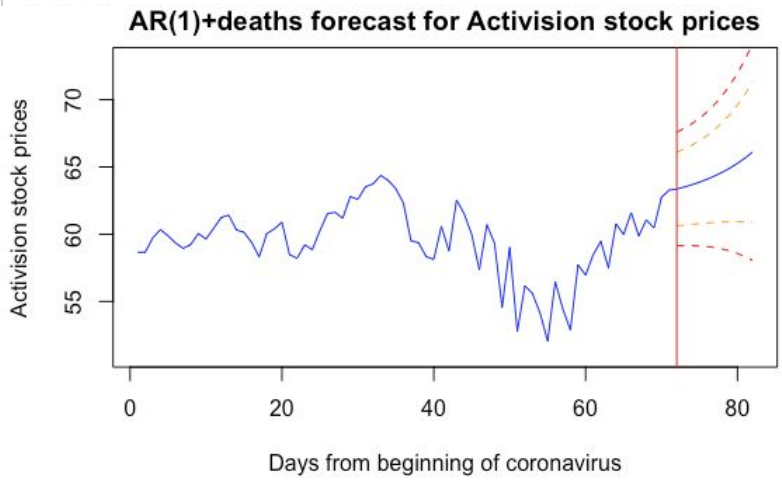
Stocks mostly tend to fall during the next 11 days from 62.3 to 59.84929

-The other model includes deaths_lags.

$$atvi_{level} = \beta_0 + \beta_1 atvi_{lags} + \beta_2 deaths_{lags} + \varepsilon$$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.607e+01  5.027e+00   3.196  0.00212 **
atvi_lags   7.289e-01  8.451e-02   8.625 1.61e-12 ***
deaths_lags 9.836e-06  9.297e-06   1.058  0.29383
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model is statistically significant, but does not explain a lot of variation. Also, $deaths_{lags}$ is not significant.



On the plot we see that stocks will slightly grow (63.35781; 66.10806) 80% prediction interval lower bound is (60.62; 60.8), upper bound (66;71) for next 11 days

95% prediction interval lower bound is (58; 59), upper bound (67;74) for next 11 days

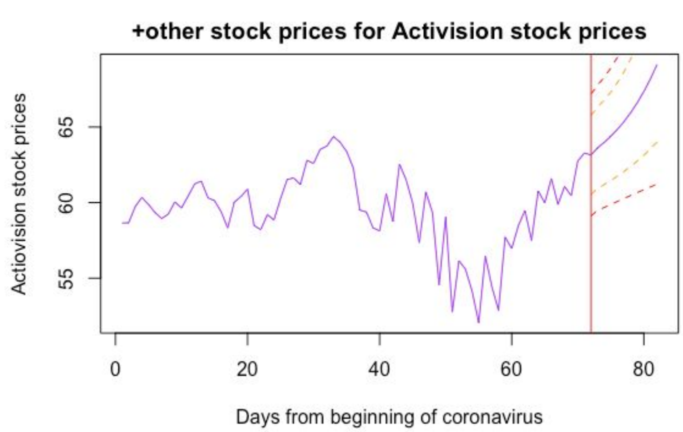
-The third AR1 model will include other companies' stocks.

$$atvi_{level} = \beta_0 + \beta_1 atvi_{lags} + \beta_2 deaths_{lags} + \beta_3 nflx\_lags + \beta_4 wmt\_lags + \beta_5 dpz\_lags + \varepsilon$$

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.625e+01  8.218e+00   4.412 3.95e-05 ***
atvi_lags    7.700e-01  1.273e-01   6.050 7.94e-08 ***
deaths_lags  2.988e-05  1.185e-05   2.522   0.0141 *
nflx_lags    1.649e-02  1.554e-02   1.061   0.2926
wmt_lags    -1.862e-01  7.584e-02  -2.456   0.0167 *
dpz_lags    -2.282e-02  8.728e-03  -2.615   0.0111 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 1.722 on 65 degrees of freedom
Multiple R-squared:  0.6039,    Adjusted R-squared:  0.5734
F-statistic: 19.82 on 5 and 65 DF,  p-value: 6.063e-12
```

All coefficients except of $nflx\_lags$ are statistically significant.
The model is also statistically significant with low p-value and RSE.



On the plot we see that stocks are increasing during the next 11 days from 63 to 69.
80% of stocks will increase on 4-9 units.

95% of stocks will increase on 2-10 units.

Walmart Ar1

- -And for **Walmart** we have similar Ar1 model with its historical stocks as predictor variable

$$wmt_{level} = \beta_0 + \beta_1 wmt_{lags} + \varepsilon$$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  41.1730    12.0834   3.407   0.0011 **
wmt_lags      0.6466     0.1040   6.215  3.4e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.216 on 69 degrees of freedom
Multiple R-squared:  0.3589,	Adjusted R-squared:  0.3496
F-statistic: 38.63 on 1 and 69 DF,  p-value: 3.4e-08
```

The model has low R-squared, but p-value is still smaller than 0.05 which means our model has significance for us. Also, RSE is pretty low.

We can notice the dramatic fall of Walmart's stocks. According to point forecast they will change from 124 to 116.
80% of stocks will decrease on 7-9 units.

95% of stocks will decrease on 6-10 units.

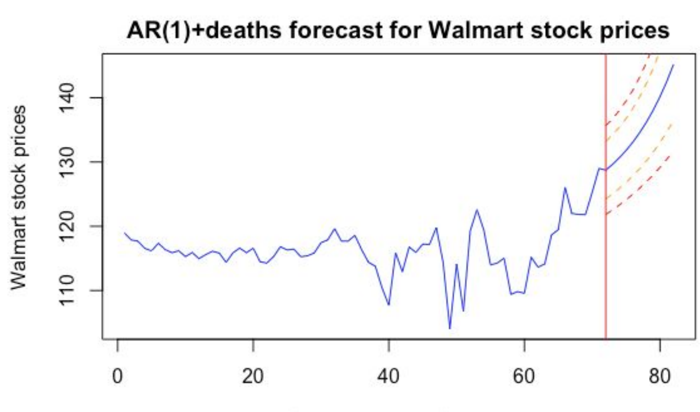
- More advanced prediction model for Walmart has deaths_lags

$$wmt_{level} = \beta_0 + \beta_1 wmt_{lags} + \beta_2 deaths_{lags} + \varepsilon$$

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.984e+01  1.241e+01   4.821 8.37e-06 ***
wmt_lags    4.798e-01  1.076e-01   4.459 3.17e-05 ***
deaths_lags 5.891e-05  1.684e-05   3.498 0.000831 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.983 on 68 degrees of freedom
Multiple R-squared:  0.4567,	Adjusted R-squared:  0.4407
F-statistic: 28.58 on 2 and 68 DF,  p-value: 9.809e-10
```

It has higher R squared, but still much smaller than 1.
P-value is much higher than in the model without deaths lag, but this model is still statistically significant.



From the plot we see that Walmart stocks will have an extreme raise. According to point forecast stocks will rise from 128.7352 to 145.2079 during the next 11 days.

80% of stocks will increase on 12-20 units.

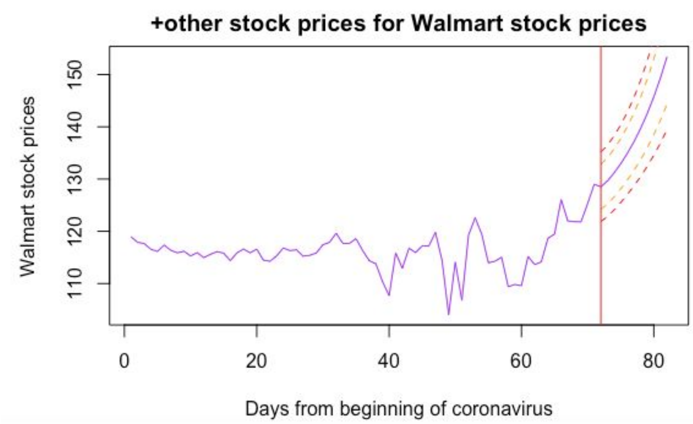
95% of stocks will increase on 10-23 units.

- Ar1 model with other stocks' lags
  $wmt_{level} = \beta_0 + \beta_1 wmt_{lags} + \beta_2 deaths_{lags} + \beta_3 atvi\_lags + \beta_4 nflx\_lags + \beta_5 dpz\_lags + \varepsilon$

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.823e+01  1.355e+01   5.774 2.36e-07 ***
wmt_lags      3.333e-01  1.250e-01   2.666  0.00968 **
deaths_lags   9.350e-05  1.953e-05   4.787 1.01e-05 ***
atvi_lags     3.149e-01  2.098e-01   1.501  0.13824
nflx_lags    -3.135e-02  2.562e-02  -1.223  0.22563
dpz_lags     -3.008e-02  1.439e-02  -2.090  0.04051 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.839 on 65 degrees of freedom
Multiple R-squared:  0.5296,    Adjusted R-squared:  0.4934
F-statistic: 14.63 on 5 and 65 DF,  p-value: 1.341e-09
```

Model has 3 statistically significant coefficients - wmt_{lags} , $deaths_{lags}$, dpz_lags
Model itself is statistically significant with p-value 1.341e-09
and a small RSE.



Stocks will have an extreme rise in the next 11 days , from 128.5230 to 153.4304.
80% of stocks will increase on 20-30 units.

95% of stocks will increase on 18-32 units.

Step 4
On this step we will check the accuracy of our prediction.We will compare the forecasted values with the actual ones and look at the errors of forecasted data.
As we had 3 prediction AR1 models per company, we will choose the best one for each company and make a conclusion based on it.

Netflix
There is a table with some type of errors of the predicted data.
Obviously, we see that the AR1 model that includes deaths lag ($nflx_{level} = \beta_0 + \beta_1 nflx_{lags} + \beta_2 deaths_{lags} + \varepsilon$) is the most relevant, because all the errors of this prediction are the smallest among other predictions.

	NFLX AR(1) <dbl>	NFLX AR+deaths <dbl>	NFLX AR+deaths+stocks <dbl>
ME	36.984421	1.6206326	2.0179094
RMSE	38.148505	10.5460351	11.2270356
MAE	36.984421	9.1756670	9.6959508
MPE	8.591323	0.3473372	0.4390547
MAPE	8.591323	2.1297059	2.2500830
MASE	3.959908	0.4959708	0.5240937

In general, the errors of this AR1 model are pretty small.
Doing a straightforward comparison, on some days prediction was really close. What is more, actual stocks fall into the 80% prediction interval.

Date	Close
15.04.2020	426.75
16.04.2020	439.170013
17.04.2020	422.959991
20.04.2020	437.48999
21.04.2020	433.829987
22.04.2020	421.420013
23.04.2020	426.700012

1	417.9539
2	420.7456
3	423.6464
4	426.9218
5	430.6202
6	434.7960
7	439.5108

Predicted stocks for model with deaths lag

Actual Nflx's stocks data

Domino's Pizza

In 2 models ME is negative, that means that the predicted value is less than the true value, but it is still close to 0, which is pretty good.

	DPZ AR(1) <dbl>	DPZ AR+deaths <dbl>	DPZ AR+deaths+stocks <dbl>
ME	19.519198	-1.7549741	-3.6200451
RMSE	23.301043	6.6660024	7.2486409
MAE	19.618515	5.8379466	6.5606797
MPE	5.247061	-0.5167158	-1.0242809
MAPE	5.275068	1.5868222	1.7914576
MASE	2.558154	0.2121493	0.2384133

ME of first model (only with company's lags) is pretty far from zero.
Comparing the errors of 3 models, we can conclude that the model $dmz_{level} = \beta_0 + \beta_1 dmz_{lags} + \beta_2\ cases_{lags} + \varepsilon$ is the best.
Comparing the actual stocks data with predicted by model above, we can say that forecasted values are really close to the real ones. Also, these values fall into 80% prediction interval.

Date	Close
15.04.2020	354.619995
16.04.2020	360.470001
17.04.2020	362.970001
20.04.2020	370.480011
21.04.2020	365.220001
22.04.2020	383.75
23.04.2020	369.640015

1	363.9377
2	365.2181
3	366.6272
4	368.1778
5	369.8841
6	371.7618
7	373.8282

Activision

	ATVI AR(1) <dbl>	ATVI AR+deaths <dbl>	ATVI AR+deaths+stocks <dbl>
ME	5.593842	2.611065	2.060253
RMSE	5.683410	2.793203	2.459967
MAE	5.593842	2.611065	2.078510
MPE	8.396789	3.909019	3.079167
MAPE	8.396789	3.909019	3.106880
MASE	3.925309	1.353308	1.077286

From the table we see that the smallest error has the model with atvi, deaths and stocks lags.

$$atvi_{level} = \beta_0 + \beta_1 atvi_{lags} + \beta_2 deaths_{lags} + \beta_3 nflx\_lags + \beta_4 wmt\_lags + \beta_5 dpz\_lags + \varepsilon$$

Date	Close
15.04.2020	65.709999
16.04.2020	68.050003
17.04.2020	66.879997
20.04.2020	66.5
21.04.2020	65.720001
22.04.2020	66.980003
23.04.2020	65.879997

63.14951
63.62648
63.98414
64.38830
64.84496
65.36094
65.94390

From straightforward comparison, we see that results are similar with small errors.

Walmart

	WMT AR(1) <dbl>	WMT AR+deaths <dbl>	WMT AR+deaths+stocks <dbl>
ME	10.728128	-1.763227	-3.085597
RMSE	11.108355	3.709952	5.414961
MAE	10.728128	2.956541	4.195267
MPE	8.219287	-1.372658	-2.391023
MAPE	8.219287	2.275079	3.231328
MASE	4.897115	1.125637	1.597254

The smallest error has a model with atvi and deaths lags: $wmt_{level} = \beta_0 + \beta_1 wmt_{lags} + \beta_2 deaths_{lags} + \varepsilon$

Date	Close
15.04.2020	128.76
16.04.2020	132.33
17.04.2020	132.12
20.04.2020	129.85001
21.04.2020	129.21001
22.04.2020	131.59
23.04.2020	128.53

128.7352
129.6393
130.6589
131.8087
133.1055
134.5679
136.2171

Empirical comparison shows also pretty similar values. Also, actual values fall into the 80% prediction interval.

Conclusion

To sum up the research, we have conducted several stages of analysis that included:

-understanding the existence and strength of relationship between stocks and other factors like Covid-19 deaths/cases and stocks of complement on-demand companies

From this step we found out that there is pretty strong dependence between those factors. Of course, stocks are such phenomena that are pretty hard to predict, but what if our research will clarify that a bit, at least in nowaday situation. So, Covid-19 obviously had a positive impact on such on-demand companies like Netflix, Domino's Pizza, Activision and Walmart. And this is not strange - being locked in our houses we all watch movies, use food delivery services, play online games and from time to time go to the biggest supermarkets to fill up our fridges.

-making the prediction of stocks of each company

We had 3 AR1 models for each company with different predictor variables.

-deciding which prediction is better

With the help of Autoregression we have made predictions that were pretty similar to actual stocks.

Main conclusion - stocks tend to grow, so it is a great opportunity for those who have on-demand companies' stocks to make money on it.

All in all, many branches of the economy of different countries have suffered from the impact of Covid-19, but there are many such companies that continue to make money and contribute to their country's wealth.