The economic impact of COVID across people in different demographic groups and education levels.

Team Members:

Chris Capps

Timothy Keating

Vedika Nigam

Abazar Rahma

Introduction

- COVID has had a huge impact on the economy and our lives. The impact of COVID has not been uniform across different groups.
- By conducting this study, we hope to examine how COVID has affected individuals in terms of their employment situation.
- Findings from this study could help policy makers in creating appropriate support structures for affected individuals.
- For this study, we have narrowed our focus to North Carolina.

Data Source

- Integrated Public Use Microdata Series (IPUMS) which is the world's largest individual-level population database.
- IPUMS has compiled this data from American Community Survey (ACS) which is a demographics survey program conducted by the U.S. Census Bureau.

Questions we hope to answer

What are the demographic and educational attainment factors that predict who is unable to work in North Carolina during COVID?

Data Exploration

Cleaning the data involved the following steps:

- Filtering the data for North Carolina by using STATE FIP

```
]:  # Filtering data for North Carolina
    demographic_data_df_NC = demographic_data_df[demographic_data_df['STATEFIP'] == 37]
    demographic_data_df_NC
```

]:

	YEAR	MONTH	STATEFIP	METAREA	COUNTY	AGE	SEX	RACE	MARST	HISPAN	EDUC	COVIDTELEW	COVIDUNAW
63258	2020	5	37	3122	0	54	1	100	6	0	111	1	1
63259	2020	5	37	3121	37067	51	2	100	4	0	91	1	1
63260	2020	5	37	3121	37067	49	1	100	6	0	111	1	1
63261	2020	5	37	1521	37119	65	2	100	1	0	73	99	1
63262	2020	5	37	1521	37119	61	1	100	1	0	73	2	1

- Removing invalid inputs (values that are 99) for target variable

```
:  # Filtering target columns to keep valid data and drop 99 values
   demographic_data_df_NC = demographic_data_df_NC[demographic_data_df_NC['COVIDUNAW'] != 99]
   demographic_data_df_NC.head(10)
```

:

	YEAR	MONTH	STATEFIP	METAREA	COUNTY	AGE	SEX	RACE	MARST	HISPAN	EDUC	COVIDTELEW	COVIDUNAW
63258	2020	5	37	3122	0	54	1	100	6	0	111	1	1
63259	2020	5	37	3121	37067	51	2	100	4	0	91	1	1
63260	2020	5	37	3121	37067	49	1	100	6	0	111	1	1

Data Exploration

- Explore data using value_counts(). Binning the independent variables using map function

```
education={111:"Bachelor's",
            73:"High School or below",
            81:"Some College or Associate Degree",
           123:"Graduate or Professional Degree",
           92:"Some College or Associate Degree",
           91:"Some College or Associate Degree",
           125:"Graduate or Professional Degree",
           60:"High School or below",
           50:"High School or below",
           124:"Graduate or Professional Degree",
           71:"High School or below",
}
```

```
#Applying map function to change categorical data from numbers to labels
demographic_data_df_NC["education"] = demographic_data_df_NC['EDUC'].map(education)
demographic_data_df_NC.head()
```

	YEAR	MONTH	METAREA	COUNTY	AGE	SEX	RACE	MARST	HISPAN	EDUC	COVIDTELEW	COVIDUNAW	gender	education
63258	2020	5	3122	0	54	1	100	6	0	111	1	1	Male	Bachelor's
63259	2020	5	3121	37067	51	2	100	4	0	91	1	1	Female	Some College or Associate Degree
63260	2020	5	3121	37067	49	1	100	6	0	111	1	1	Male	Bachelor's
63262	2020	5	1521	37119	61	1	100	1	0	73	2	1	Male	High School or below
63268	2020	5	3122	0	35	2	100	1	0	111	2	1	Female	Bachelor's

Data Exploration

- Binning Age variable using pd.cut

```
demographic_data_df_NC['age'] = pd.cut(x=demographic_data_df_NC['AGE'], bins=[16, 24, 34, 44, 54, 64,90],
                    labels=['16 to 24', '25 to 34', '35 to 44',
                            '45 to 54', '55 to 64','65+'])
```

```
demographic_data_df_NC
```

MONTH	METAREA	COUNTY	AGE	SEX	RACE	MARST	HISPAN	EDUC	COVIDTELEW	COVIDUNAW	gender	education	race	hispanic	marital_status	age
5	3122	0	54	1	100	6	0	111	1	1	Male	Bachelor's	White	Non-Hispanic	Single	45 to 54

- Finally removing all Nan values and exporting clean file to csv

```
#Dropping all Nan values
demographic_data_df_NC.dropna(how='any',inplace = True)
demographic_data_df_NC
```

	YEAR	MONTH	METAREA	COUNTY	AGE	SEX	RACE	MARST	HISPAN	EDUC	COVIDTELEW	COVIDUNAW	gender	education	race	hispanic
63258	2020	5	3122	0	54	1	100	6	0	111	1	1	Male	Bachelor's	White	Non-Hispanic

Data Exploration

Total Data Points: 19205 individuals who answered the survey

Independent variables (features):

- Age - 16-24, 25-34, 35-44, 45-54, 55-64, 65+
- Gender - Male, Female
- Race - Black, White, Native American, Asian
- Marital Status - Married, Single, Divorced
- Education- High school or below, Some College, Bachelor's, Graduate or Professional Degree

Dependent variable (target):

- COVIDUNAW - individuals who are unable to work during COVID (1 : able to work, 2: unable to work)

Data Exploration

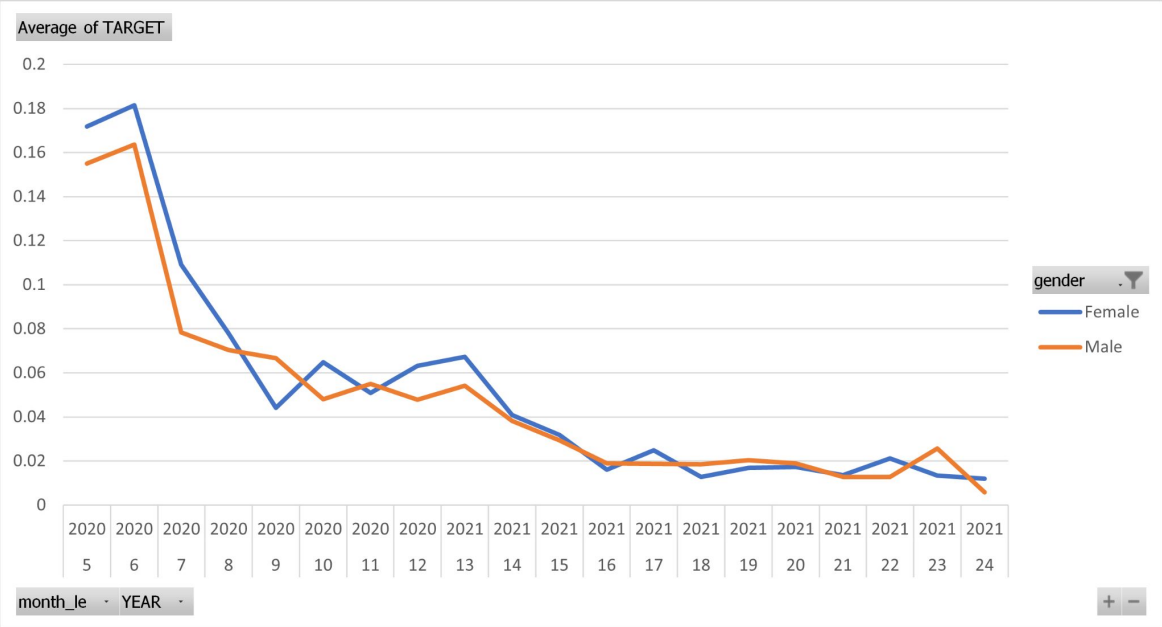
Statistics summary using describe()

```
1 df_encoded.describe()
```

	AGE	COVIDUNAW	month_le	gender_Female	gender_Male	education_Bachelor's	education_Graduate or Professional Degree	education_High School or below	education_Some College or Associate Degree
count	19205.000000	19205.000000	19205.000000	19205.000000	19205.000000	19205.000000	19205.000000	19205.000000	19205.000000
mean	43.509399	1.045040	15.103306	0.477636	0.522364	0.275397	0.156313	0.308774	0.259516
std	14.770140	0.207398	5.575414	0.499513	0.499513	0.446726	0.363162	0.462000	0.438380
min	16.000000	1.000000	5.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	31.000000	1.000000	10.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	43.000000	1.000000	15.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000
75%	55.000000	1.000000	20.000000	1.000000	1.000000	1.000000	0.000000	1.000000	1.000000
max	85.000000	2.000000	24.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

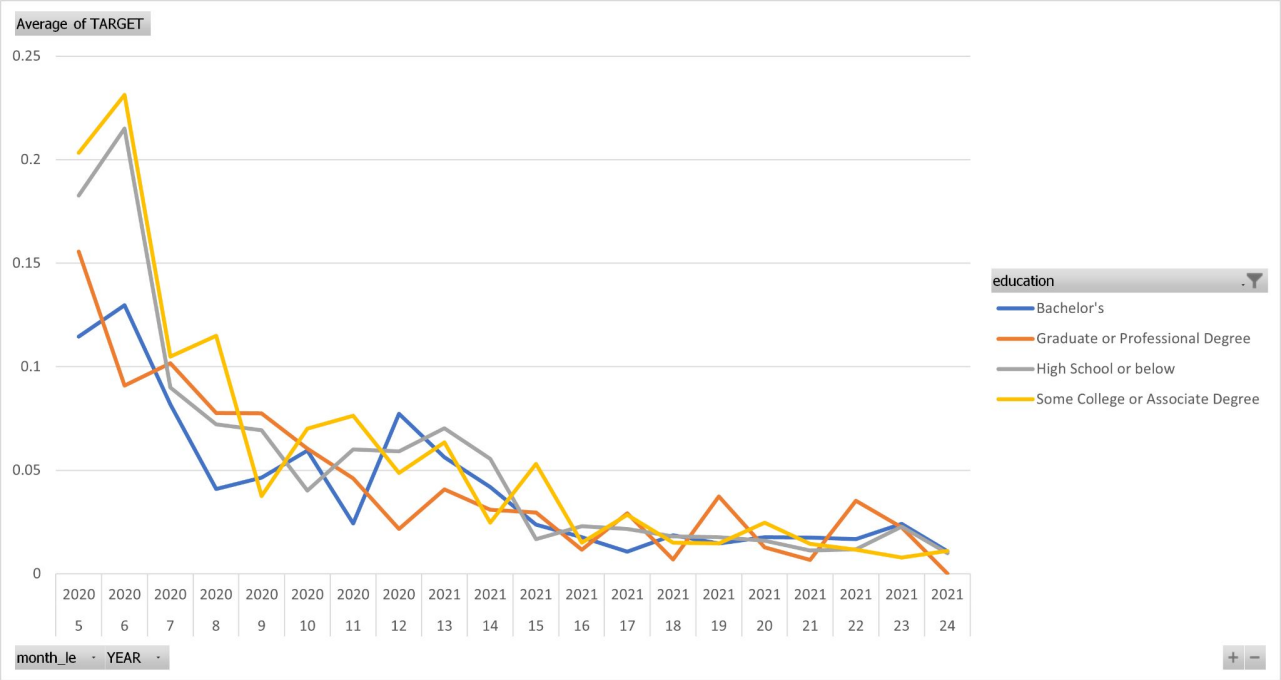
Data Analysis - Gender

Women were affected more in the beginning but over time unemployment due to COVID declined to similar and low levels for both genders



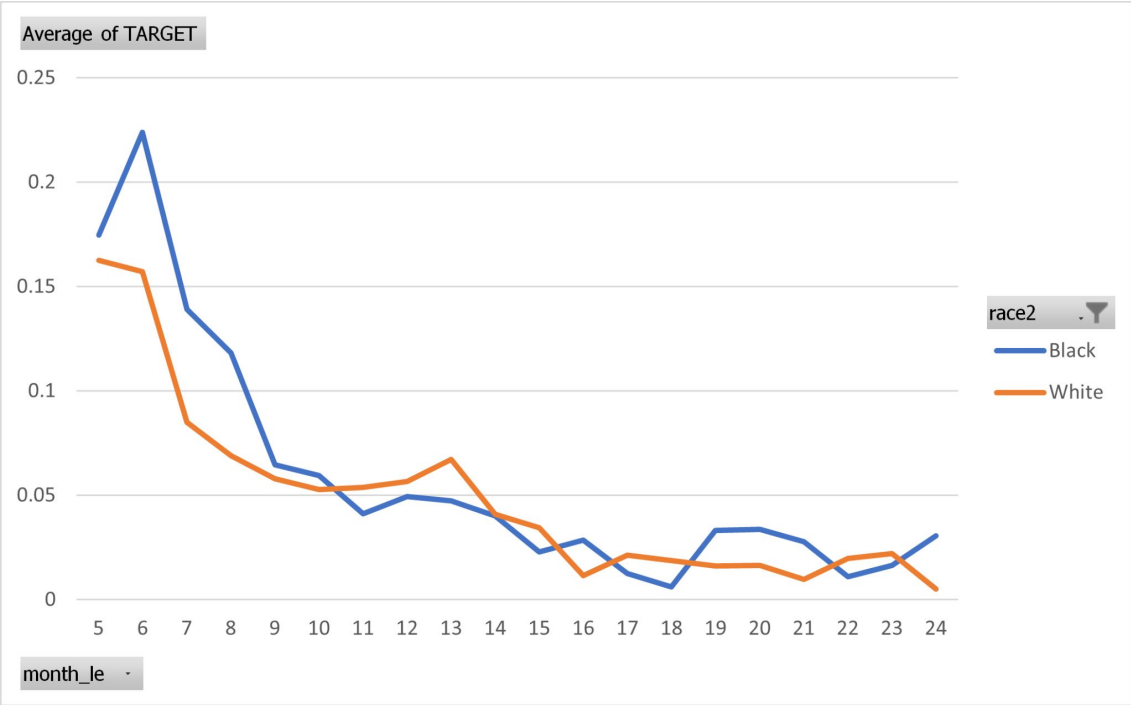
Data Analysis - Education

Individuals without a bachelor degree were affected more in the beginning but by December 2021 unemployment due to COVID declined for all categories



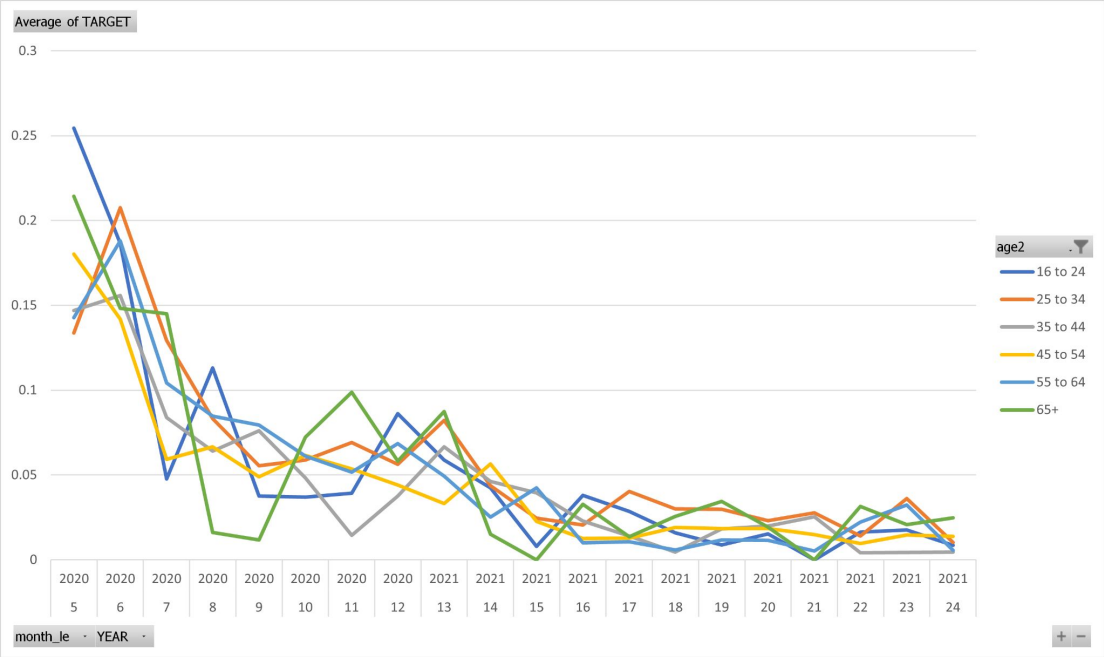
Data Analysis - Race

Blacks were affected more during first several months of COVID though overtime rates declined at a somewhat similar level for both black and whites.



Data Analysis - Age

There is no clear pattern across ages though at different times in the pandemic different categories were worst affected.



Dashboard

The dashboard will be created in Tableau - the link to which will be embedded in a web application. The interactive element will be the map of North Carolina with categorical variables as layers. The following sheets will be created as part of the Dashboard. The demographic characteristics and education sheets will include data analysis including statistics and line graphs. The data modeling sheet will include analysis from the machine learning model.