

Assignment 5

by Alisiya Balayan, Bishoy Abdelmalik, Arian
Dehghani, Ariel Kohanim, Natalie Weingart,
Sergio Ramirez





TABLE OF CONTENTS

1. Datasets

3. Feature Selection

2. Datasets Attributes

4. Approach



Our Datasets

- **Covid-19 World Vaccination Progress**
 - This dataset tells us about: which country is using what vaccine, which country's vaccination programme is more advanced, and where the rate of vaccinated people per day is higher in terms of percent from the entire population.
- **Novel Coronavirus 2019 Dataset**
 - This dataset has daily level information on the number of affected cases, deaths and recovery from COVID-19. It provides us with data about each country and their cases, deaths and recoveries.
- **Countries population by year 2020**
 - This dataset provides us with the world population and top 20 countries' live clock. It contains population data for the past, present and future.
- **Covid-19 Daily Vaccination**
 - This data set contains vaccination data for countries showing how many people are vaccinated daily.



Types of Features

- Numeric
- Categorical
- Time-series



Feature Selection

- We have selected several numeric features from all datasets as well as categorical and time-series.
- Numeric data will be used to calculate if our hypothesis is true while categorical and time-series will measure the spread of vaccinations over time per different states.



Approach

- Numeric input and output:
 - Pearson algorithm
- Categorical input and output:
 - ANOVA regression analysis
- Time-series analysis:
 - Cross-correlation test



Upcoming..

- Once we selected all of the features and have strong reasoning behind it, we can proceed to build our predictive models.

