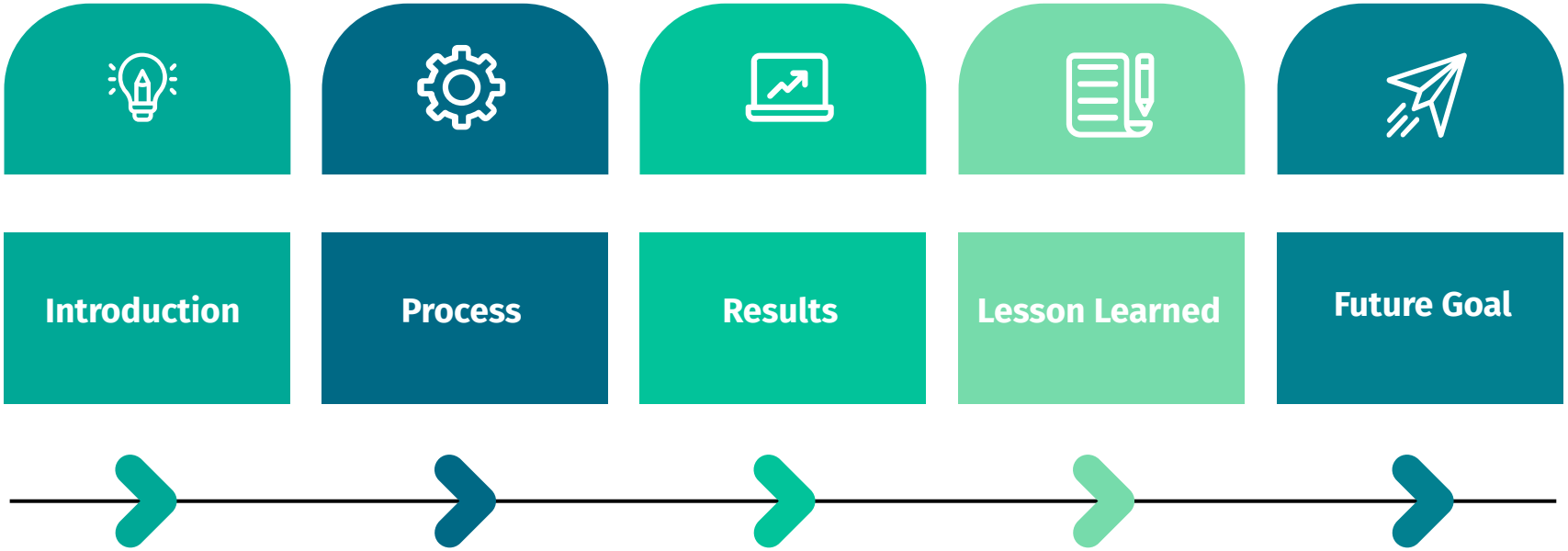
Job Market & GDP Growth Analysis

Presented By:

Oluyomi Alabi

Tina Pham

Agenda



Introduction

- Volatile labor market from the pandemic.
- Create a stats model to analyze a **job market dataset** of top 58 global economies from 2000-2016.
- <u>Questions:</u>
  - How socio-economic affect unemployment rate?
  - How does the GDP growth correlates with the proportion of labor force with basic education?
  - Are there any significant difference in GDP Growth and the educational attainment of the labor force?

Tools

Job market data of **top 58 economies** in the world

- Jupyter Notebook
- Pandas

- Seaborn
- Matplotlib
- Tableau

- Statsmodels
- Scikit Learn

Process

1. **Obtain the data**
   World Bank Data _Job data set
2. **Data Cleaning**
   Structure data for analysis
3. **EDA**
   Visualization to understand data
4. **Build Stat model**
   Prediction model

Unemployment Analysis

Plot of Overall Unemployment and GDP overtime

Synchronization with similar pattern in the line graph overtime.

Unemployment Analysis

OLS Regression Results

Dep. Variable:	Unemployment_total_	R-squared:	0.252
Model:	OLS	Adj. R-squared:	0.245
Method:	Least Squares	F-statistic:	36.57
Date:	Fri, 14 Jul 2023	Prob (F-statistic):	4.76e-56
Time:	12:10:22	Log-Likelihood:	-2761.9
No. Observations:	986	AIC:	5544.
Df Residuals:	976	BIC:	5593.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	14.6381	0.572	25.582	0.000	13.515	15.761
Labor_force_total_	-5.291e-09	1.2e-09	-4.392	0.000	-7.66e-09	-2.93e-09
Agriculture_value_added_%GDP_	-0.2388	0.024	-9.889	0.000	-0.286	-0.191
GDP_growth_	-0.1148	0.032	-3.582	0.000	-0.178	-0.052
GDP_per_capita_	-7.686e-05	1.16e-05	-6.635	0.000	-9.96e-05	-5.41e-05
GDP_per_person_employed_	-1.273e-05	7.12e-06	-1.788	0.074	-2.67e-05	1.24e-06
Manufacturing_value_added_%GDP	-0.0708	0.022	-3.279	0.001	-0.113	-0.028
Inflation_consumer_prices_	-0.0212	0.024	-0.897	0.370	-0.068	0.025
Exports_of_goods_and_services_%GDP_	-0.0200	0.005	-4.419	0.000	-0.029	-0.011
High-technology_exports_	-0.0518	0.012	-4.320	0.000	-0.075	-0.028

...

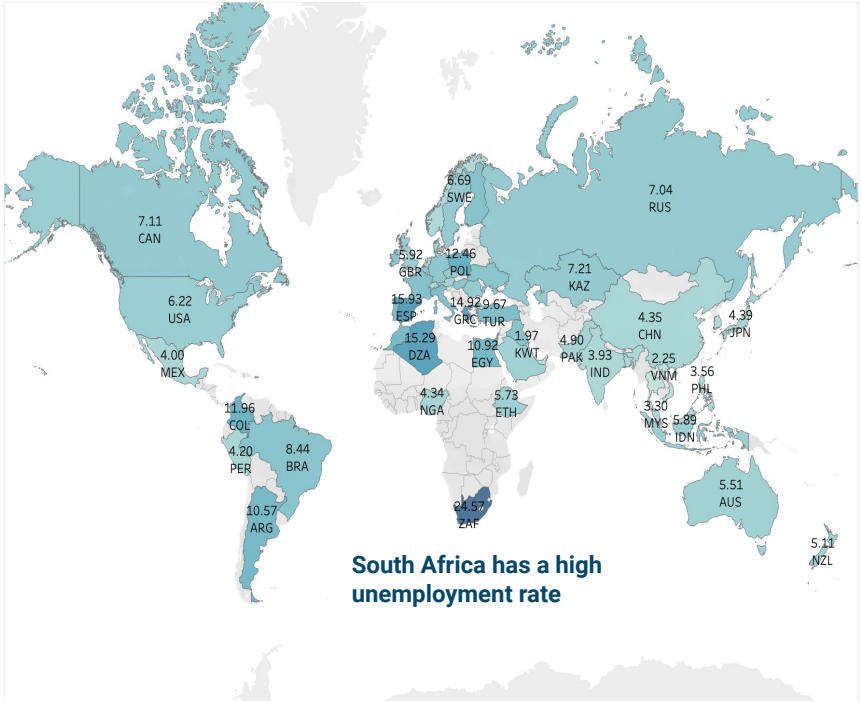
Linear Regression Model:

- **Small Adj. R-squared**
  - Model only explain 25% of population.

→ Mix of type of economy within the population.

- **p-value for GDP < 0.05:**
  - p-value for GDP growth is smaller than 0.05 → dependency
  - Negatively correlate to unemployment
- **p-value for Inflation >0.05:**
  - Relationship due to natural variation.
  - Possible outliers within the population → more cleaning

Unemployment Analysis

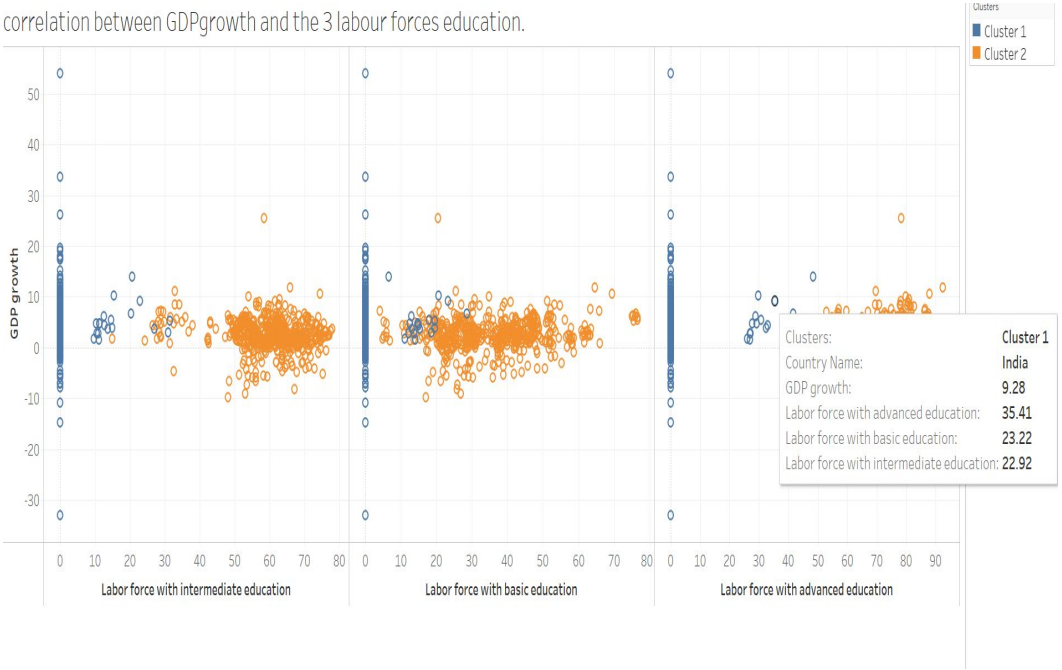


Map of unemployment rate around the world:

→ Variation in type of economy across the population could reflect on small Adj. R-squared

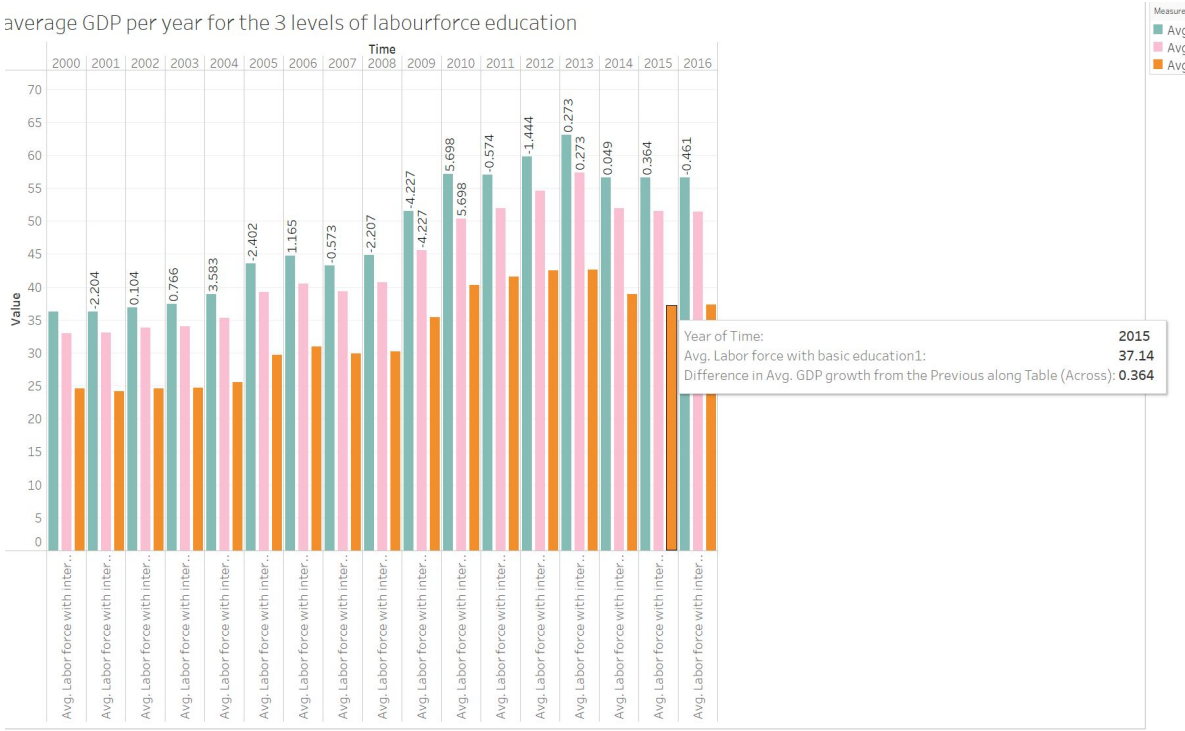
→ Need **more cleaning** in the data with possible **detailed analysis** to train the model.

Scattered plots showing correlation between GDP and Labour Forces Education attainment.



This result show the relationship between the dataset. It indicates potential pattern or trend in the data. As GDP growth increases, some labour force with advance education reduces but more of basic education increases showing negative and positive correlation.

Grouped -bar charts showing GDP growth difference between the 3 levels of Education in workforce.



There are no significant difference in GDP growth on the educational attainment of the labour force per year.

Regression model

OLS Regression Results

Dep. Variable:	GDP_growth_	R-squared:	0.630
Model:	OLS	Adj. R-squared:	0.561
Method:	Least Squares	F-statistic:	9.084
Date:	Thu, 13 Jul 2023	Prob (F-statistic):	8.00e-06
Time:	23:12:16	Log-Likelihood:	-62.663
No. Observations:	39	AIC:	139.3
Df Residuals:	32	BIC:	151.0
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.2158	0.812	2.729	0.010	0.562	3.870
Private_credit_bureau_coverage_	-0.0055	0.006	-0.885	0.383	-0.018	0.007
Total_employment,_total_	7.152e-07	1.42e-07	5.052	0.000	4.27e-07	1e-06
Labor_force_with_basic_education_	0.0259	0.019	1.366	0.182	-0.013	0.064
Labor_force,_total_	-7.833e-07	1.4e-07	-5.583	0.000	-1.07e-06	-4.97e-07
Population_ages_0-14,_total_	1.876e-07	5.39e-08	3.479	0.001	7.78e-08	2.97e-07
Population_ages_65_and_above,_total_	8.517e-08	8.13e-08	1.048	0.302	-8.04e-08	2.51e-07

Omnibus:	4.429	Durbin-Watson:	2.465
Prob(Omnibus):	0.109	Jarque-Bera (JB):	3.547
Skew:	0.735	Prob(JB):	0.170
Kurtosis:	3.155	Cond. No.	2.58e+08

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

The R-squared value which is the coefficient of determination shows the proportion of the variance in the dependent variable(GDP growth) that is predictable by the independent variables.
R-squared indicate that approximately 63% of the variance in the dependent variable(GDP growth) that can be predicted by the independent variables included in the linear regression model.
It shows this model fits but there is still room for adjustment by removing independent variables whose p-value >0.05

Lesson Learned

1

Organization

Put in time for organization.

From the data model to the repo

2

Collaboration

Communication.

Documentation.

Learn from each other.

3

Data

This type of data project has no finish line.

4

Syntax

WHITESPACE!!!!!

5

Data Cleaning

😰😱😭

Never ending

Future Goal

Enhance

- Introduce more variables and more current data to enhance a stats modeling.

Cleaning

- Perform a more refined analysis on unemployment on different country to train regression model.

Build

- Build a machine learning model for resource allocation.

Thank You