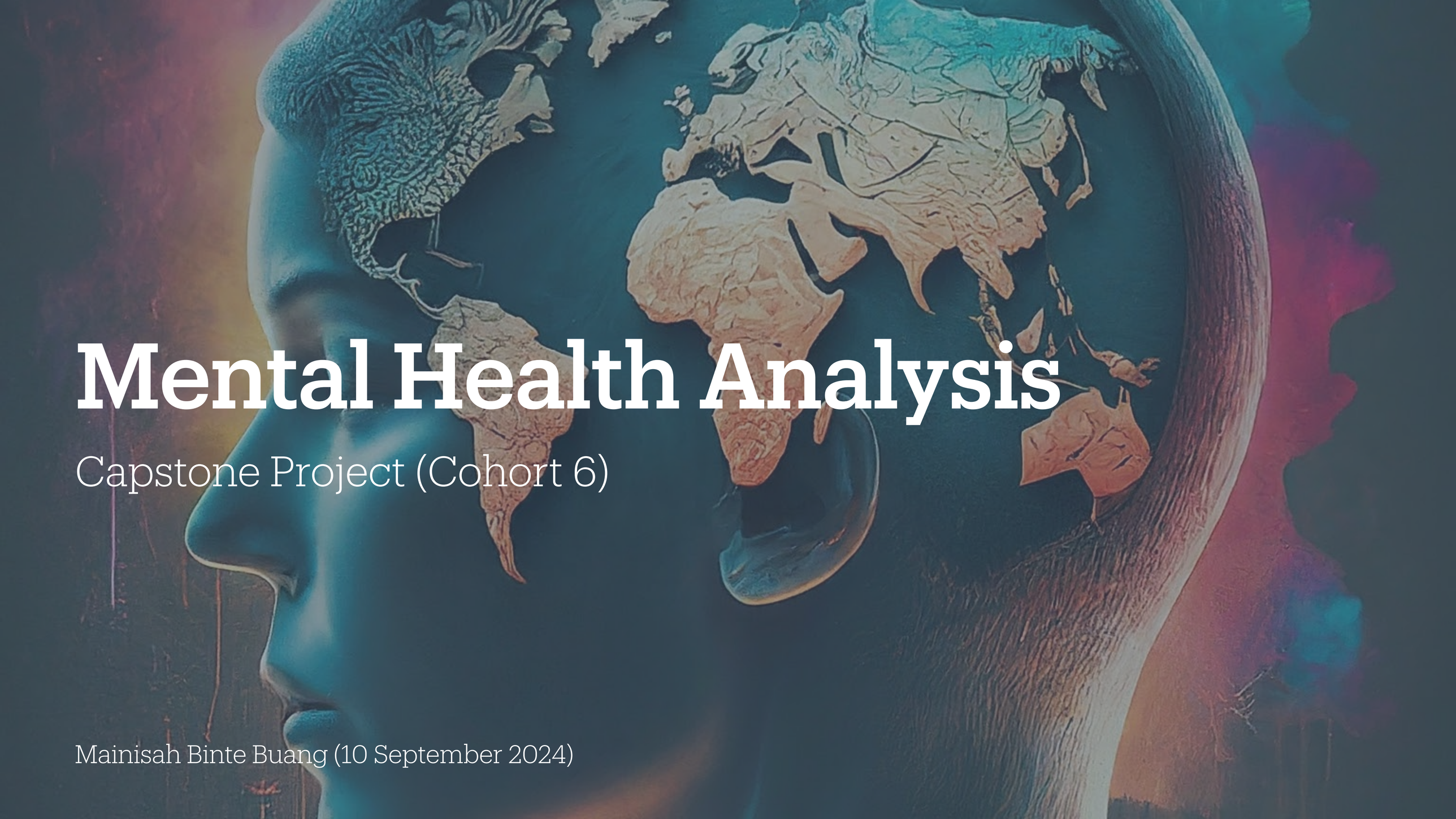


Mental Health Analysis

Capstone Project (Cohort 6)

Mainisah Binte Buang (10 September 2024)

1.Unveiling the Stress Epidemic:

A Data-Driven Exploration of Mental Health

- Mental health is integral to overall well-being, impacting how we think, feel, and act.
- Stress, a prevalent mental health challenge, affects millions globally.
- This project delves into stress trends, contributing factors, and potential solutions using data analysis and machine learning.

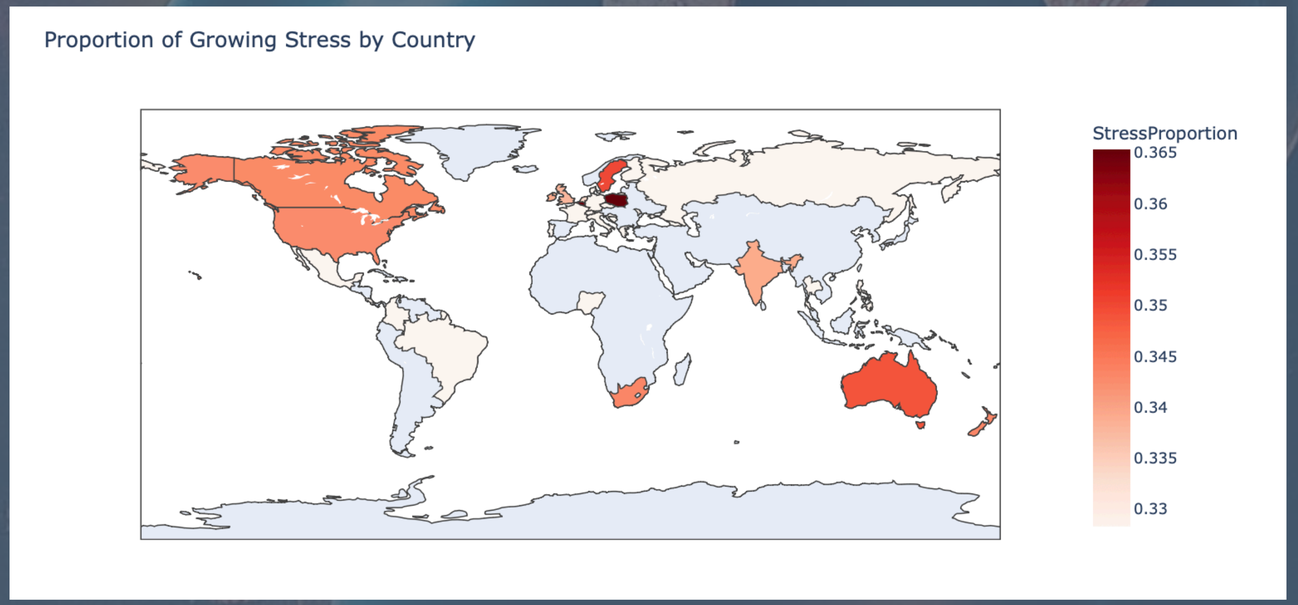
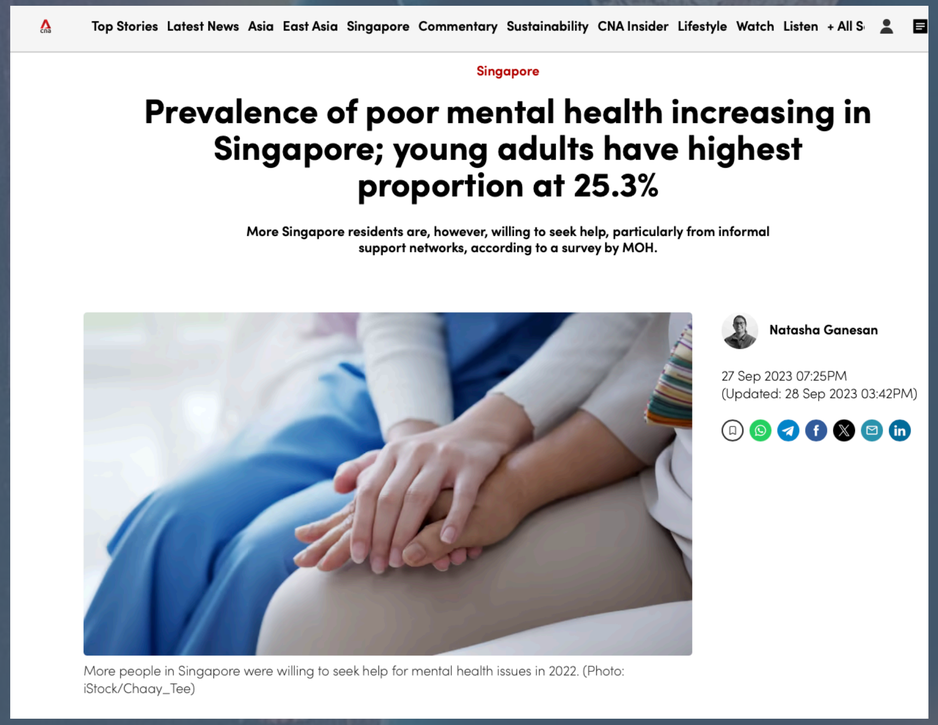


Image generated by Gemini

2. Why Stress Matters

- Stress is a widespread phenomenon with significant impacts on individuals and society.
- Understanding stress trends and contributing factors is crucial for promoting mental health.
- The analysis will focus on identifying countries and demographics most affected by stress.



Top Stories Latest News Asia East Asia Singapore Commentary Sustainability CNA Insider Lifestyle Watch Listen + All S

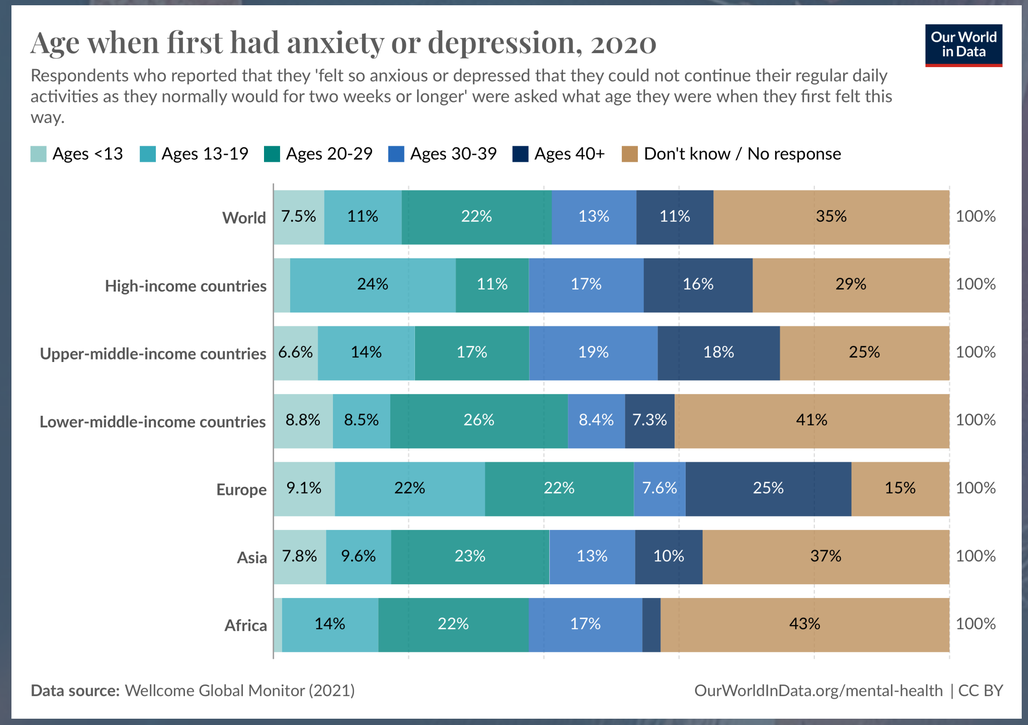
Singapore

Prevalence of poor mental health increasing in Singapore; young adults have highest proportion at 25.3%

More Singapore residents are, however, willing to seek help, particularly from informal support networks, according to a survey by MOH.

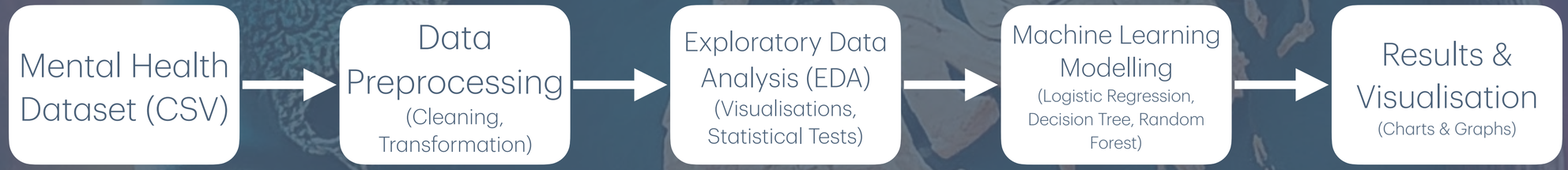
More people in Singapore were willing to seek help for mental health issues in 2022. (Photo: iStock/Chaay_Tee)

https://www.channelnewsasia.com/singapore/poor-mental-health-young-adults-seek-help-moh-survey-3802531



https://ourworldindata.org/mental-health

3. System Overview & Objectives



A. Find the country with the biggest increase in stress:

1. Relative change: Captures the percentage increase in stress, providing insights into the rate of change.
2. Absolute change: Indicates the actual numerical increase in stress levels, offering a concrete measure of the change.
3. Prevalence: Highlights the current proportion of the population experiencing stress, providing context for the increase.

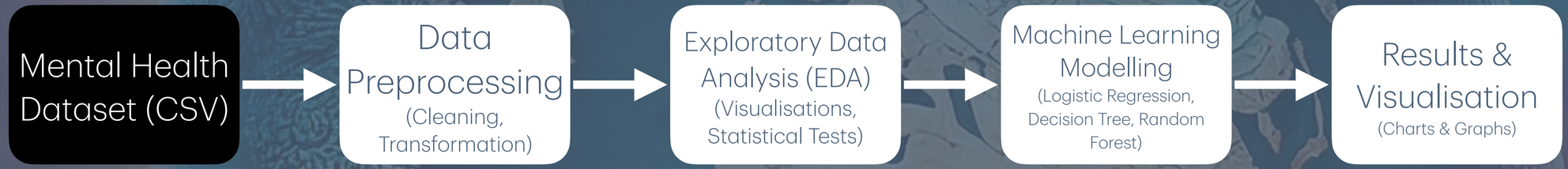
B. Understand the stress trends in these countries.

C. Figure out why stress is rising in Singapore and suggest solutions.

D. Identify countries with lower stress that might be better places to live.

E. Suggestions on how to handle and reduce stress.

4. Codebase Highlights



This dataset records a global survey conducted to track trends in mental health. The data covers a range of variables such as levels of stress, depression, anxiety, subjective well-being, and use of mental health services. The survey involved respondents from various demographic backgrounds, including gender, employment status, and geographic region. This dataset aims to provide a better understanding of changes in mental health globally over the specified time period.

https://www.kaggle.com/datasets/divaniazzahra/mental-health-dataset

```
(292364, 17)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 292364 entries, 0 to 292363
Data columns (total 17 columns):
 #   Column                  Non-Null Count   Dtype 
---  ------                  --------------   ----- 
 0   Timestamp               292364 non-null  object
 1   Gender                  292364 non-null  object
 2   Country                 292364 non-null  object
 3   Occupation              292364 non-null  object
 4   self_employed           287162 non-null  object
 5   family_history          292364 non-null  object
 6   treatment               292364 non-null  object
 7   Days_Indoors            292364 non-null  object
 8   Growing_Stress          292364 non-null  object
 9   Changes_Habits          292364 non-null  object
 10  Mental_Health_History   292364 non-null  object
 11  Mood_Swings             292364 non-null  object
 12  Coping_Struggles        292364 non-null  object
 13  Work_Interest           292364 non-null  object
 14  Social_Weakness         292364 non-null  object
 15  mental_health_interview 292364 non-null  object
 16  care_options            292364 non-null  object
dtypes: object(17)
memory usage: 37.9+ MB
None
```

	Timestamp	Gender	Country	Occupation	self_employed	family_history	treatment	Days_Indoors	Growing_Stress	Changes_Habits	Mental_Health_History	Mood_Swings	Coping_Struggles	Work_Interest	Social_Weakness	mental_health_interview	care_options
0	2014-08-27 11:29:31	Female	United States	Corporate	NaN	No	Yes	1-14 days	Yes	No	Yes	Medium	No	No	Yes	No	Not sure
1	2014-08-27 11:31:50	Female	United States	Corporate	NaN	Yes	Yes	1-14 days	Yes	No	Yes	Medium	No	No	Yes	No	No
2	2014-08-27 11:32:39	Female	United States	Corporate	NaN	Yes	Yes	1-14 days	Yes	No	Yes	Medium	No	No	Yes	No	Yes
3	2014-08-27 11:37:59	Female	United States	Corporate	No	Yes	Yes	1-14 days	Yes	No	Yes	Medium	No	No	Yes	Maybe	Yes
4	2014-08-27 11:43:36	Female	United States	Corporate	No	Yes	Yes	1-14 days	Yes	No	Yes	Medium	No	No	Yes	No	Yes

4. Codebase Highlights

Mental Health Dataset (CSV) → **Data Preprocessing** (Cleaning, Transformation) → Exploratory Data Analysis (EDA) (Visualisations, Statistical Tests) → Machine Learning Modelling (Logistic Regression, Decision Tree, Random Forest) → Results & Visualisation (Charts & Graphs)

- # Fill missing values in `self_employed` with 'Unknown
  df['self_employed'].fillna('Unknown', inplace=True)
- df.describe().T
- df.head()

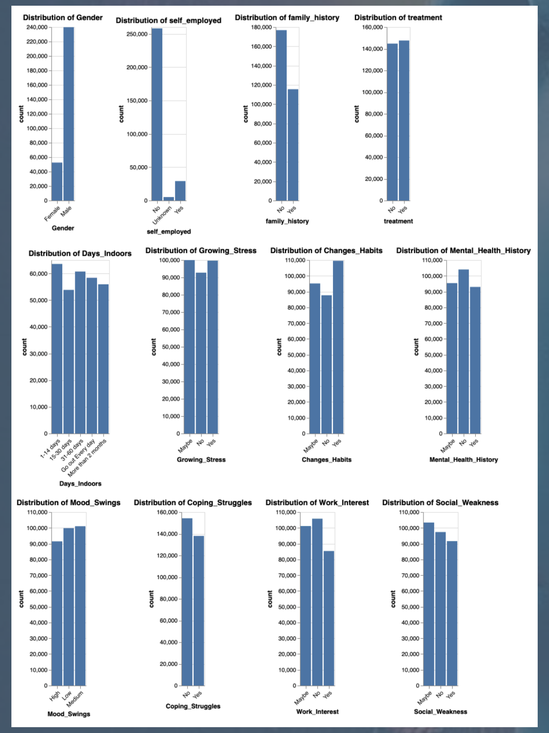
	count	unique	top	freq
Timestamp	292364	734	2014-08-27 12:31:41	780
Gender	292364	2	Male	239850
Country	292364	35	United States	171308
Occupation	292364	5	Housewife	66351
self_employed	292364	3	No	257994
family_history	292364	2	No	176832
treatment	292364	2	Yes	147606
Days_Indoors	292364	5	1-14 days	63548
Growing_Stress	292364	3	Maybe	99985
Changes_Habits	292364	3	Yes	109523
Mental_Health_History	292364	3	No	104018
Mood_Swings	292364	3	Medium	101064
Coping_Struggles	292364	2	No	154328
Work_Interest	292364	3	No	105843
Social_Weakness	292364	3	Maybe	103393
mental_health_interview	292364	3	No	232166
care_options	292364	3	No	118886

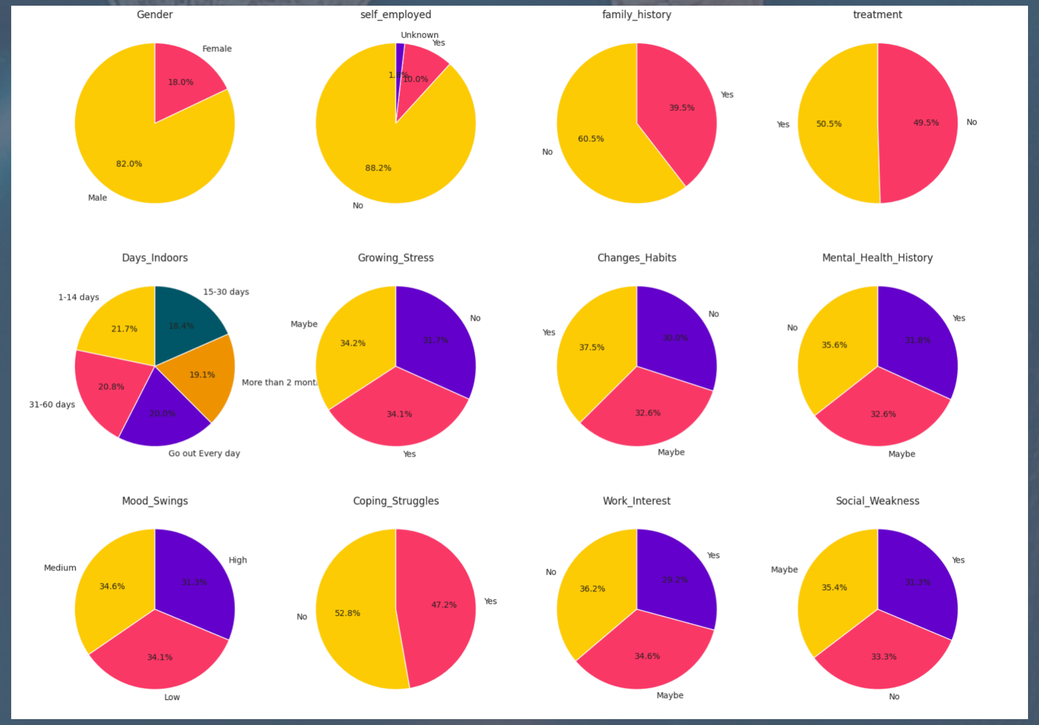
	Timestamp	Gender	Country	Occupation	self_employed	family_history	treatment	Days_Indoors	Growing_Stress	Changes_Habits	Mental_Health_History	Mood_Swings	Coping_Struggles	Work_Interest	Social_Weakness	mental_health_interview	care_options
0	2014-08-27 11:29:31	Female	United States	Corporate	Unknown	No	Yes	1-14 days	Yes	No	Yes	Medium	No	No	Yes	No	Not sure
1	2014-08-27 11:31:50	Female	United States	Corporate	Unknown	Yes	Yes	1-14 days	Yes	No	Yes	Medium	No	No	Yes	No	No
2	2014-08-27 11:32:39	Female	United States	Corporate	Unknown	Yes	Yes	1-14 days	Yes	No	Yes	Medium	No	No	Yes	No	Yes
3	2014-08-27 11:37:59	Female	United States	Corporate	No	Yes	Yes	1-14 days	Yes	No	Yes	Medium	No	No	Yes	Maybe	Yes
4	2014-08-27 11:43:36	Female	United States	Corporate	No	Yes	Yes	1-14 days	Yes	No	Yes	Medium	No	No	Yes	No	Yes

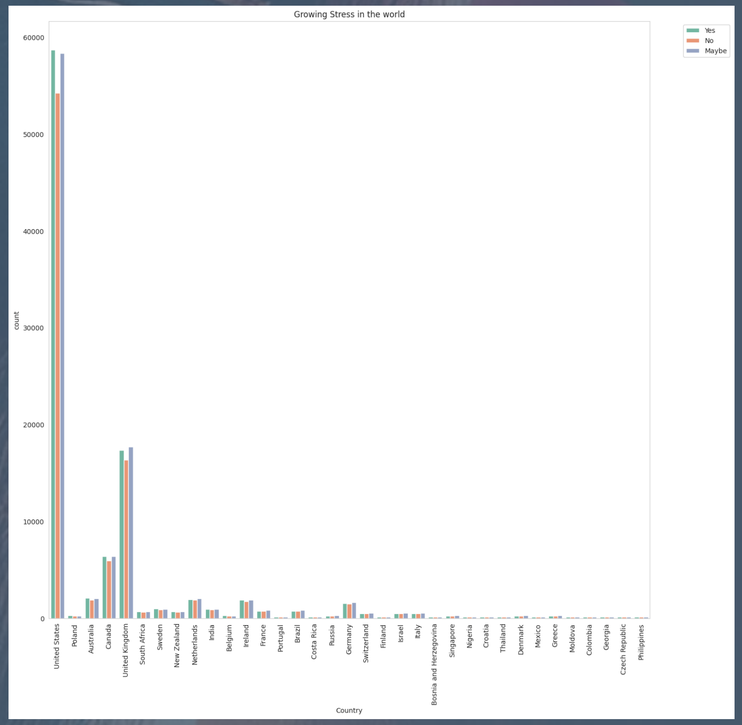
5. Model Performance

Mental Health Dataset (CSV) → Data Preprocessing (Cleaning, Transformation) → **Exploratory Data Analysis (EDA) (Visualisations, Statistical Tests)** → Machine Learning Modelling (Logistic Regression, Decision Tree, Random Forest) → Results & Visualisation (Charts & Graphs)

- Use visualisations like bar charts and pie charts to illustrate relationships and trends.







5. Model Performance

Mental Health Dataset (CSV) → Data Preprocessing (Cleaning, Transformation) → **Exploratory Data Analysis (EDA) (Visualisations, Statistical Tests)** → Machine Learning Modelling (Logistic Regression, Decision Tree, Random Forest) → Results & Visualisation (Charts & Graphs)

A. Find the country with the biggest increase in stress (Relative change, Absolute change and Prevalence)

We chose those specific columns **'Growing_Stress', 'Changes_Habits', 'Coping_Struggles', 'Work_Interest', 'Social_Weakness')** as stress-related variables for the analysis primarily because they directly reflect the presence or absence of stress-related symptoms or behaviors.

Top 3 Countries by Stress Growth:

Current Dataset (Prevalence):

	Country
1	Belgium
25	Poland
30	Sweden

Relative Change:

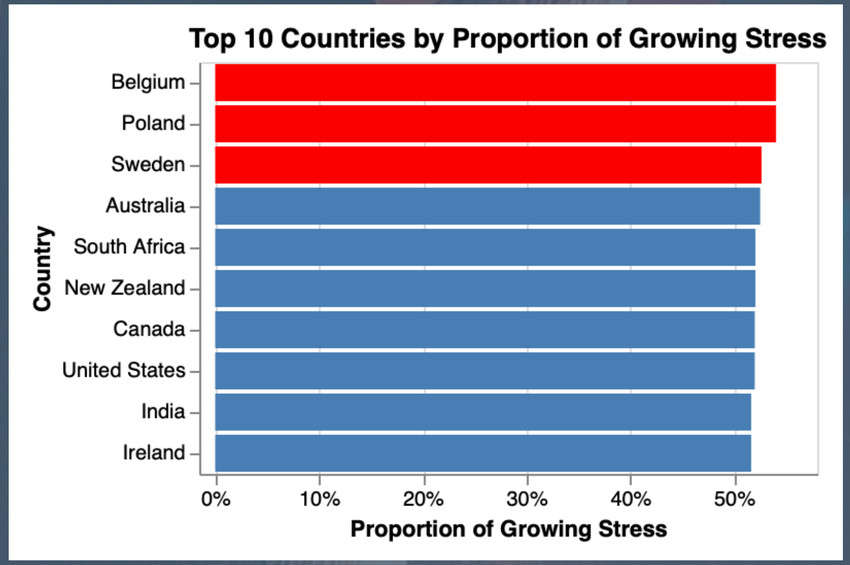
	Country
0	New Zealand
1	Belgium
2	Israel

Absolute Change:

	Country
0	New Zealand
1	Belgium
2	Israel

Stress Score for all Countries:

Country	StressProportion
Belgium	0.54
Poland	0.54
Sweden	0.526
Australia	0.525
South Africa	0.521
New Zealand	0.521
United States	0.52
Canada	0.52
India	0.516
Ireland	0.516



5. Model Performance

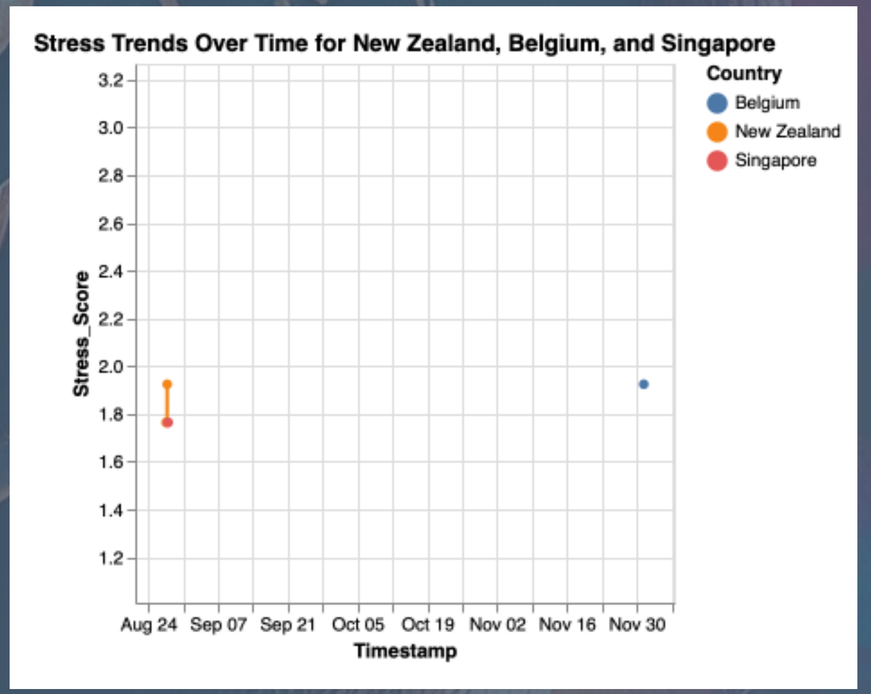
Mental Health Dataset (CSV) → Data Preprocessing (Cleaning, Transformation) → **Exploratory Data Analysis (EDA) (Visualisations, Statistical Tests)** → Machine Learning Modelling (Logistic Regression, Decision Tree, Random Forest) → Results & Visualisation (Charts & Graphs)

Comparison and Analysis of Stress Trends

- Belgium: While having the highest prevalence of growing stress based on the current dataset, the graph reveals that Belgium's stress levels remained relatively stable over time, with a slight increase towards the end of the observed period. This suggests that although a larger proportion of individuals in Belgium report experiencing growing stress, the overall increase in stress levels might not be as dramatic

- New Zealand: Identified as the country with the highest relative and absolute stress growth, the graph demonstrates a clear upward trend in stress scores over time. This indicates a substantial increase in stress levels compared to the initial state, supporting its top ranking in both relative and absolute change metrics.

- Singapore: The graph shows a relatively flat trend for Singapore, suggesting that stress levels remained fairly constant throughout the observed period. This aligns with its lower ranking in both relative and absolute stress growth metrics.

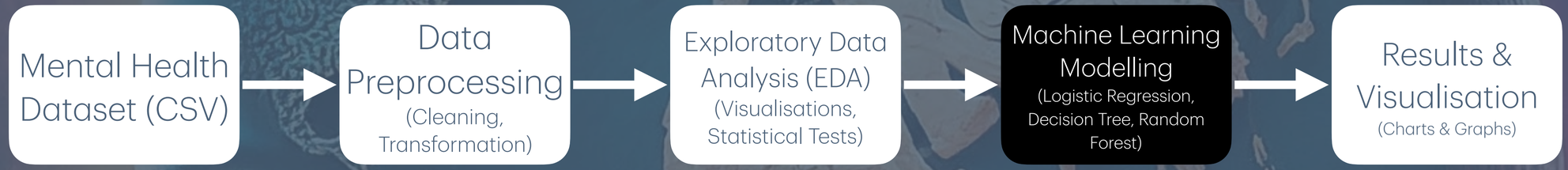


Stress Growth Analysis:

-Relative Change: New Zealand shows the highest relative increase in stress levels over time (16.09%).

-Absolute Change: New Zealand also exhibits the highest absolute increase in stress levels (0.16).

5. Model Performance



- **Inconsistency** in the timestamps across different countries can indeed introduce **bias** and make the comparison of stress trends **less reliable**. Using the timestamp for relative and absolute stress growth calculations might not be the most suitable approach in this case

- Instead, we can leverage other variables or features in the dataset to create a more meaningful and fair comparison. One potential approach is to focus on **the prevalence of growing stress** across different countries, as this metric is not dependent on the timestamp and provides a snapshot of the current situation in each country.

- We can also explore the association between Growing_Stress' and other variables like Gender, self_employed, family_history, treatment, Mental_Health_History, and Mood_Swings using statistical tests or predictive models. This can help **identify potential risk factors** and provide insights into the underlying causes of growing stress, regardless of the temporal aspect.

5. Model Performance

Mental Health Dataset (CSV) → Data Preprocessing (Cleaning, Transformation) → Exploratory Data Analysis (EDA) (Visualisations, Statistical Tests) → **Machine Learning Modelling** (Logistic Regression, Decision Tree, Random Forest) → Results & Visualisation (Charts & Graphs)

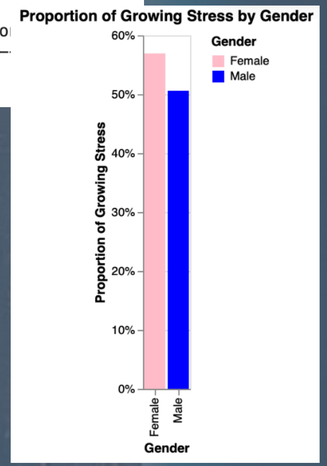
B. Understand the stress trends in these countries.

Prevalence of Growing Stress:

- By Gender: Females report a slightly higher proportion of growing stress (56.9%) compared to males (50.6%).
- By Country: Belgium and Poland have the highest proportion of respondents reporting growing stress (both at 54%).
- By Occupation: Individuals in 'Business' and 'Student' roles report the highest proportion of growing stress (57.1% and 56.7%, respectively).

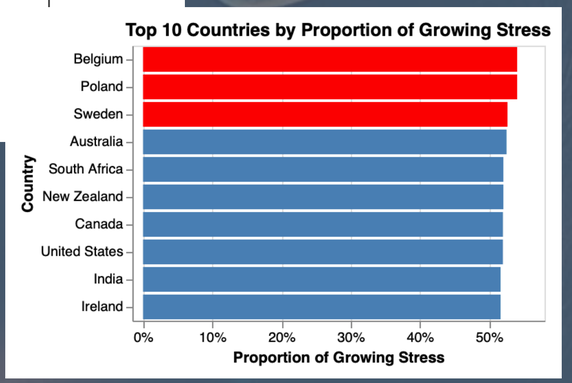
Proportion of 'Growing_Stress' by Gender:

Gender	StressProportion
Female	0.569079
Male	0.505929



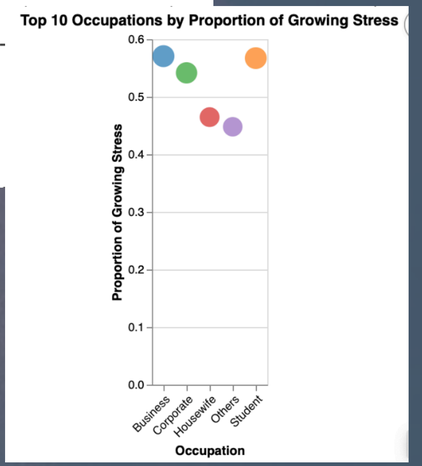
Top 10 Countries by Proportion of 'Growing_Stress':

Country	StressProportion
Belgium	0.540395
Poland	0.540395
Sweden	0.526428
Australia	0.525131
South Africa	0.520517
New Zealand	0.520517
Canada	0.519919
United States	0.519863
India	0.516465
Ireland	0.516465



Top 10 Occupations by Proportion of 'Growing_Stress':

Occupation	StressProportion
Business	0.570514
Student	0.567084
Corporate	0.541535
Housewife	0.464631
Others	0.447896



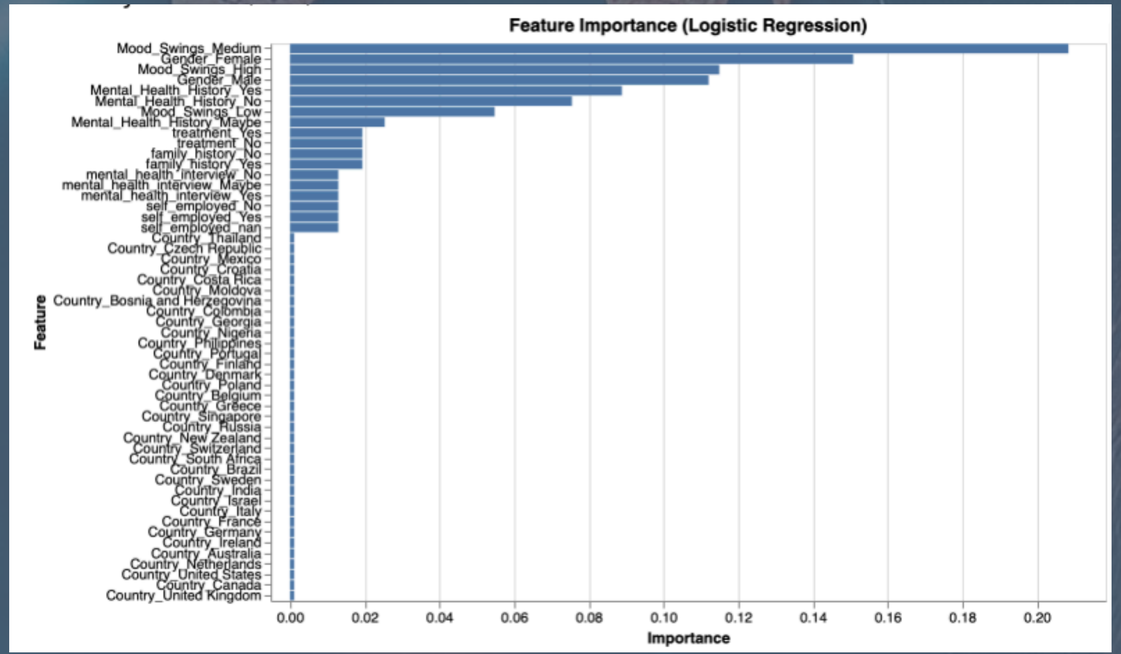
5. Model Performance

Mental Health Dataset (CSV) → Data Preprocessing (Cleaning, Transformation) → Exploratory Data Analysis (EDA) (Visualisations, Statistical Tests) → **Machine Learning Modelling** (Logistic Regression, Decision Tree, Random Forest) → Results & Visualisation (Charts & Graphs)

Predictive Modeling:

- Logistic Regression: Achieves an accuracy of 55.4%. The most influential predictors are Mood_Swings_Medium and Gender_Female (odds ratio 1.16).
- Decision Tree Classifier: Achieves an accuracy of 53.4%.

```
Logistic Regression Model Evaluation:

Accuracy: 0.553046054683439

Confusion Matrix:
 [[ 9063  9616]
 [ 7581 12216]]

Classification Report:
              precision    recall  f1-score   support

         0.0       0.54      0.49      0.51     18679
         1.0       0.56      0.62      0.59     19797

    accuracy                           0.55     38476
   macro avg       0.55      0.55      0.55     38476
weighted avg       0.55      0.55      0.55     38476


Odds Ratios:

| Predictor                      | Odds Ratio   |
|:-------------------------------|:-------------|
| Gender_Female                  | 1.16274      |
| Gender_Male                    | 0.893982     |
| Country_Australia              | 1.00111      |
| Country_Belgium                | 1.00111      |
| Country_Bosnia and Herzegovina | 1.00111      |
| Country_Brazil                 | 1.00111      |
| Country_Canada                 | 1.00111      |
| Country_Colombia               | 1.00111      |
| Country_Costa Rica             | 1.00111      |
| Country_Croatia                | 1.00111      |
```




```
Decision Tree Classifier Model Evaluation:

Accuracy: 0.5344890321239214

Confusion Matrix:
 [[ 8079 10600]
 [ 7311 12486]]

Classification Report:
              precision    recall  f1-score   support

         0.0       0.52      0.43      0.47     18679
         1.0       0.54      0.63      0.58     19797

    accuracy                           0.53     38476
   macro avg       0.53      0.53      0.53     38476
weighted avg       0.53      0.53      0.53     38476
```

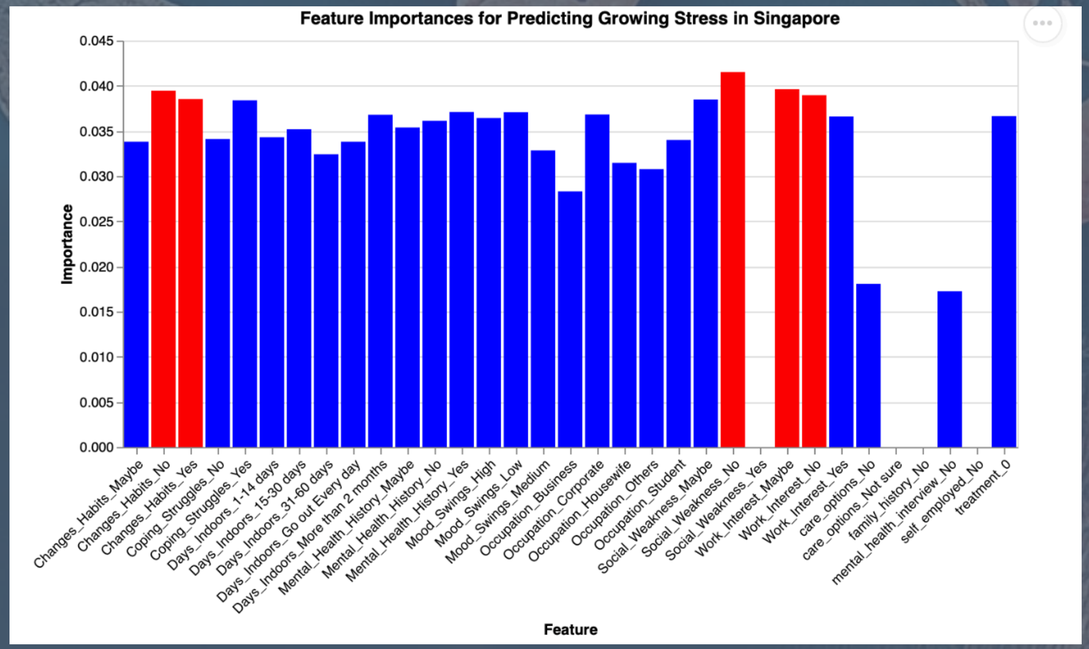
The accuracy values of around 50% for the Logistic Regression and Decision Tree models in analysis indicate that **these models are performing poorly**. In essence, they are barely doing better than random guessing in predicting the 'Growing_Stress' outcome.

6. Key Insights

Mental Health Dataset (CSV) → Data Preprocessing (Cleaning, Transformation) → Exploratory Data Analysis (EDA) (Visualisations, Statistical Tests) → **Machine Learning Modelling (Logistic Regression, Decision Tree, Random Forest)** → Results & Visualisation (Charts & Graphs)

C. Figure out why stress is rising in Singapore and suggest solution

```
Decision Tree Results:
Accuracy: 0.8333333333333334
              precision    recall  f1-score   support

       Maybe       0.78      0.87      0.82        53
          No       0.89      0.77      0.82        43
         Yes       0.85      0.85      0.85        60

    accuracy                           0.83       156
   macro avg       0.84      0.83      0.83       156
weighted avg       0.84      0.83      0.83       156

[[46  1  6]
 [ 7 33  3]
 [ 6  3 51]]

Random Forest Results:
Accuracy: 0.7628205128205128
              precision    recall  f1-score   support

       Maybe       0.77      0.81      0.79        53
          No       0.69      0.72      0.70        43
         Yes       0.82      0.75      0.78        60

    accuracy                           0.76       156
   macro avg       0.76      0.76      0.76       156
weighted avg       0.77      0.76      0.76       156

[[43  4  6]
 [ 8 31  4]
 [ 5 10 45]]
```



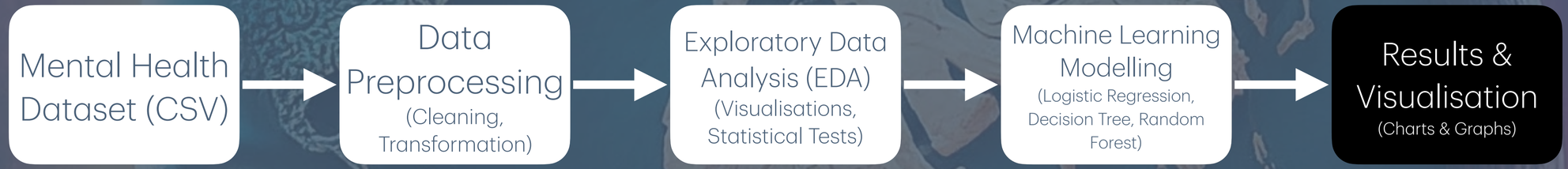
Overall, the chart visualizes the importance of various features in predicting "Growing Stress" in Singapore, as determined by a Random Forest model.

Top 5 Important Features (in red):

1.Changes_Habits_Yes
2.Changes_Habits_No
3.Days_Indoors_More than 2 month
4.Mood_Swings_High
5.Occupation_Corporate

Decision Tree Classifier: Achieves an accuracy of 83.3%
Random Forest : Achieves an accuracy of 76.28%.

6. Key Insights



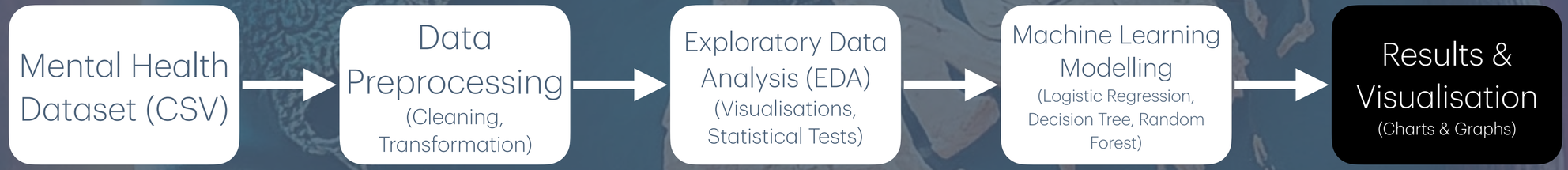
Recommendations:

- Promote Healthy Coping Mechanisms: Implement programs that equip individuals with the skills to manage changes in habits, regulate mood swings, and cope with stress effectively
- Encourage Outdoor Activity & Social Connection: Advocate for policies and initiatives that promote outdoor spaces, physical activity, and social interaction to counterbalance the effects of prolonged indoor time.
- Workplace Wellness: Encourage organizations, particularly in the corporate sector, to prioritize employee well-being through flexible work arrangements, mental health resources, and stress management programs.
- Mental Health Awareness & Support: Increase awareness and accessibility of mental health services to address underlying emotional and psychological challenges contributing to stress.

Policy Recommendations:

- Workplace Policies: Mandatory annual leave and flexible working arrangements. Stricter regulations on overtime hours and promoting a culture of work-life balance. Mental health support programs and resources within workplaces.
- Social Safety Nets: Strengthened unemployment benefits and social assistance programs, similar to Denmark's unemployment protection, to reduce financial anxieties. Affordable housing initiatives and accessible mental health services.
- Education System: Reduced emphasis on academic competition and standardized testing. Increased focus on holistic development and social-emotional learning.
- Community Building: Investment in community spaces and programs that foster social connections and support networks.

6. Key Insights



D. Identify countries with lower stress that might be better places to live.

Filter out countries with fewer than 50 respondents to ensure adequate representation. Calculate the proportion of individuals reporting growing stress for each remaining country with the lowest proportions.

Top countries with lowest growing stress:

Country	0
Israel	0.328205
Philippines	0.328205
Nigeria	0.328205
Singapore	0.328205
Moldova	0.328205
Mexico	0.328205
Italy	0.328205
Greece	0.328205
Russia	0.328205
Georgia	0.328205
France	0.328205
Germany	0.328205
Denmark	0.328205
Czech Republic	0.328205
Croatia	0.328205
Costa Rica	0.328205
Colombia	0.328205
Switzerland	0.328205
Brazil	0.328205
Bosnia and Herzegovina	0.328205
Thailand	0.328205
Finland	0.328205
Portugal	0.328205
Netherlands	0.33339
United Kingdom	0.337717
India	0.339221
Ireland	0.339221
United States	0.342833
Canada	0.342892
New Zealand	0.343531
South Africa	0.343531
Australia	0.34849
Sweden	0.349894
Belgium	0.365291
Poland	0.365291

E. Suggestions on how to handle and reduce stress.

Mindfulness-Based Interventions: Mindfulness meditation, yoga, and breathing exercises have been shown to effectively reduce stress and improve mental well-being. Recommend accessible resources or programs for individuals to learn and practice these techniques.

Cognitive Behavioral Therapy (CBT): CBT helps individuals identify and challenge negative thought patterns and develop healthy coping mechanisms. Suggest seeking professional help or accessing online CBT resources.

Exercise and Physical Activity: Encourage regular physical activity as it releases endorphins, improves mood, and reduces stress. Recommend finding enjoyable activities and incorporating them into daily routines.

Social Connection: Emphasize the importance of strong social relationships and support networks. Encourage participation in social activities, clubs, or support groups.

Time Management and Prioritization: Help individuals develop effective time management skills to reduce feelings of overwhelm and improve work-life balance.

7. Overcoming Challenges

- Topic selection
  Railway systems in Singapore - Singapore government is too secretive.
  International datas are available. I was more interested in railway systems in Japan and Hong Kong but unfortunately the datas for these best countries in railway systems are not easily accessible.

- Where to start? - Approached trainer and batch mate for help ; lagging behind. Anxiety kicks in!

- Should I get gemini for help? - Yes! But I wasn't sure what do I want to achieve.
  Too many variables and possibilities = not so good. Stay on track with the objectives and work on it!

- Was I stress? I couldn't remember and understand the codes. Keep practising and keep learning.

- Time management is important. Be objective and stay focus. Calm down!

8. Key Takeaways and Future Directions

Key Takeaways:

- Identified key insights into stress trends and contributing factors
- Serve as a reminder of the critical importance of addressing mental health
- Creating a more supportive and resilient society that prioritises mental well being

Current dataset:

-Feature Engineeering: Create new features or transform existing ones to better capture the relationship with the outcome

-Try different algorithms: Explore more complex models like Random Forests, Gradient Boosting, or Neural Networks that can handle non-linear relationships and potentially improve accuracy.

-Address class imbalance: If present, use techniques like oversampling, undersampling, or adjusting class weights during model training.

-Hyperparameter tuning: Optimize the model's parameters to improve its performance.

Future Directions:

-Incorporating additional data sources

-Exploring more advanced machine learning techniques such as deep learning or natural language processing

-Conduct longitudinal studies to track stress trends over time would provide valuable insights

Thank You!