IFN619 - DATA ANALYTICS FOR STRATEGIC DECISION MAKERS

ASSIGNMENT 1 - PART B

QUESTION 1. How does the frequency of mental health illness and attitudes towards mental health vary by geographic location, and what are the strongest predictors of mental health illness and specific attitudes towards mental health in the workplace?

Significance of the Analysis

Mental health illness is the key to personal well-being and to function effectively in the community. The elevating number of mental health issues all round the world is an issue of major concern, and thus there is a need to develop better understanding of the factors and other varibales such as the effect of location or type of workplace or gender or family history or perception of people around. Thus, the above question focuses on the rate of occurence of mental health illness and attitude towards it based on geographical location, also attitude towards it in workplace and the highly efficient predictor of these issues in workplace. To address this question, a dataset from a 2014 survey will be used, which had questions related to our business concern.

Stakeholders

The knowledge and the understanding generated in this analysis is beneficial to the following people:

- Employers of organizations
- Individuals in the workforce
- Policy makers
- Health practitioners
- Carers
- Ministry of Health
- Healthcare Analysts

DATA

The data that will be used for the analysis is collected from a 2014 survey, with 1259 respondents answering questions related to mental health illness, giving their personal information, talking about people's perception, workplace response to the illness and more. The link to the data can be found here (https://www.kaggle.com/osmi/mental-health-in-tech-survey). It is used from a highly credible source, Kaggle, which is trusted by many data science experts. The questions included in the survey will help us in drawing insights for our business concern.

Import the required libraries.

In [1]:

```
# import the required Python libraries to process the data
import pandas as pd
import numpy as np

#for visualization
import seaborn as sns
import matplotlib.pyplot as plt
!pip install folium
import folium
```

```
WARNING: pip is being invoked by an old script wrapper. This will fa
il in a future version of pip.
Please see https://github.com/pypa/pip/issues/5599 for advice on fix
ing the underlying issue.
To avoid this problem you can invoke Python with '-m pip' instead of
running pip directly.
Collecting folium
  Using cached folium-0.10.1-py2.py3-none-any.whl (91 kB)
Requirement already satisfied: jinja2>=2.9 in /opt/conda/lib/python
3.7/site-packages (from folium) (2.11.0)
Requirement already satisfied: requests in /opt/conda/lib/python3.7/
site-packages (from folium) (2.22.0)
Requirement already satisfied: numpy in /opt/conda/lib/python3.7/sit
e-packages (from folium) (1.18.1)
Collecting branca>=0.3.0
  Using cached branca-0.4.0-py3-none-any.whl (25 kB)
Requirement already satisfied: MarkupSafe>=0.23 in /opt/conda/lib/py
thon3.7/site-packages (from jinja2>=2.9->folium) (1.1.1)
Requirement already satisfied: idna<2.9,>=2.5 in /opt/conda/lib/pyth
on3.7/site-packages (from requests->folium) (2.8)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /opt/conda/l
ib/python3.7/site-packages (from requests->folium) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.2
1.1 in /opt/conda/lib/python3.7/site-packages (from requests->foliu
m) (1.25.7)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/
python3.7/site-packages (from requests->folium) (2019.11.28)
Requirement already satisfied: six in /opt/conda/lib/python3.7/site-
packages (from branca>=0.3.0->folium) (1.14.0)
Installing collected packages: branca, folium
Successfully installed branca-0.4.0 folium-0.10.1
```

LOAD THE DATA

We upload the csv file of the datset, read it and take a quick look at it with the code below. The link for teh dataset is provided above.

In [2]:

```
# specify the location and the filenae of your dataset
filename = "survey.csv"

# load the csv dataset to a pandas DataFrame
data = pd.read_csv( filename )

# take a look at the dataset
data
```

Out[2]:

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	wo
0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes	
1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No	
2	2014-08-27 11:29:44	32	Male	Canada	NaN	NaN	No	No	
3	2014-08-27 11:29:46	31	Male	United Kingdom	NaN	NaN	Yes	Yes	
4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No	
...	...	...	...	...	...	...	...	...	
1254	2015-09-12 11:17:21	26	male	United Kingdom	NaN	No	No	Yes	
1255	2015-09-26 01:07:35	32	Male	United States	IL	No	Yes	Yes	
1256	2015-11-07 12:36:58	34	male	United States	CA	No	Yes	Yes	
1257	2015-11-30 21:25:06	46	f	United States	NC	No	No	No	
1258	2016-02-01 23:04:31	25	Male	United States	IL	No	Yes	Yes	

1259 rows × 27 columns

The dataset is from the 2014 survey on mental health illness. It has various columns ranging from the country of the respondent, age to its workplace details.

In [3]:

```
# get the dimensions of your dataset
dimensions = data.shape

print( 'General size of the dataset: ' + str( dimensions ))

# extract the number of rows and columns from your data
num_rows = dimensions[0]
num_col  = dimensions[1]
print('The dataset has ' + str( num_rows ) + ' rows and ' + str( num_col ) + ' c
olumns!' )
```

```
General size of the dataset: (1259, 27)
The dataset has 1259 rows and 27 columns!
```

Each row represents unique responses to the survey. The 27 columns that we get from the above code represent the columns with column name as it can be seen in the output.

In [4]:

```
# extract the variables of this dataset
vars = data.columns.tolist()
vars
```

Out[4]:

```
['Timestamp',
 'Age',
 'Gender',
 'Country',
 'state',
 'self_employed',
 'family_history',
 'treatment',
 'work_interfere',
 'no_employees',
 'remote_work',
 'tech_company',
 'benefits',
 'care_options',
 'wellness_program',
 'seek_help',
 'anonymity',
 'leave',
 'mental_health_consequence',
 'phys_health_consequence',
 'coworkers',
 'supervisor',
 'mental_health_interview',
 'phys_health_interview',
 'mental_vs_physical',
 'obs_consequence',
 'comments']
```

In [ ]:

```

```

CLEAN/PREPROCESS DATA

This dataset contains the responses of individuals for the 2014 survey for mental health illness. It will, to a large extent, help us in answering the above question with its variables like country, treatment, mental health consequnce, leave, benefits, to name amongst the many. The limiting factor of this dataset is that it has some unacceptable age values.

Before we begin our analysis, it is crucial to look for missing values in the dataset. Absence of data from our dataset can and cannot affect of our analysis depending on the usage of the variable to solve our purpose.

Check for missing values and data quality

In [5]:

```
# check if there are missing values:
data.info()
# there are no missing values to addresss in this dataset

# this also gives the datatypes of every column
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1259 entries, 0 to 1258
Data columns (total 27 columns):
Timestamp                    1259 non-null object
Age                          1259 non-null int64
Gender                       1259 non-null object
Country                      1259 non-null object
state                        744 non-null object
self_employed                1241 non-null object
family_history               1259 non-null object
treatment                    1259 non-null object
work_interfere               995 non-null object
no_employees                 1259 non-null object
remote_work                  1259 non-null object
tech_company                 1259 non-null object
benefits                     1259 non-null object
care_options                 1259 non-null object
wellness_program             1259 non-null object
seek_help                    1259 non-null object
anonymity                    1259 non-null object
leave                        1259 non-null object
mental_health_consequence    1259 non-null object
phys_health_consequence      1259 non-null object
coworkers                    1259 non-null object
supervisor                   1259 non-null object
mental_health_interview      1259 non-null object
phys_health_interview        1259 non-null object
mental_vs_physical           1259 non-null object
obs_consequence              1259 non-null object
comments                     164 non-null object
dtypes: int64(1), object(26)
memory usage: 265.7+ KB
```

As we can see above, there are no missing values "to address" in this dataset.

We remove the columns "Timestamp" and "comments" from our dataset "data" as part of the data cleaning process.

In [6]:

```
# Dropping columns "Timestamp" and "comments"
data.drop(['Timestamp', 'comments'], axis = 1, inplace = True)
print(data.columns)
```

```
Index(['Age', 'Gender', 'Country', 'state', 'self_employed', 'family
_history',
       'treatment', 'work_interfere', 'no_employees', 'remote_work',
       'tech_company', 'benefits', 'care_options', 'wellness_progra
m',
       'seek_help', 'anonymity', 'leave', 'mental_health_consequenc
e',
       'phys_health_consequence', 'coworkers', 'supervisor',
       'mental_health_interview', 'phys_health_interview',
       'mental_vs_physical', 'obs_consequence'],
      dtype='object')
```

ANALYSIS AND VISUALIZATION OF THE DATA

ANALYSIS OF THE DATA FOR 'FREQUENCY OF MENTAL HEALTH ILLNESS BY GEOGRAPHIC LOCATION'

Let us examine the data to find the number of people suffering from mental issues in each country. Later we find the percentage of each country to get a better understanding.

In [8]:

```
#count number of mental health illness cases in each country
cases = data['Country'].value_counts()
cases
```

Out[8]:

```
United States             751
United Kingdom            185
Canada                     72
Germany                    45
Netherlands                27
Ireland                    27
Australia                  21
France                     13
India                      10
New Zealand                 8
Sweden                      7
Switzerland                 7
Poland                      7
Italy                       7
Belgium                     6
South Africa                6
Brazil                      6
Israel                      5
Bulgaria                    4
Singapore                   4
Austria                     3
Mexico                      3
Russia                      3
Finland                     3
Greece                      2
Denmark                     2
Portugal                    2
Colombia                    2
Croatia                     2
Moldova                     1
Japan                       1
Romania                     1
Costa Rica                  1
Philippines                 1
Bosnia and Herzegovina      1
Czech Republic              1
Zimbabwe                    1
Latvia                      1
Spain                       1
Bahamas, The                1
Uruguay                     1
Norway                      1
Nigeria                     1
Hungary                     1
Thailand                    1
China                       1
Georgia                     1
Slovenia                    1
Name: Country, dtype: int64
```

In [9]:

```
#Find the total number of cases all round the world
totalcases= cases.sum()
totalcases
```

Out[9]:

```
1259
```

In [10]:

```
#calculate cases in each country by %
cases_by_percent = cases/totalcases * 100
print(cases_by_percent)
```

```
United States             59.650516
United Kingdom            14.694202
Canada                     5.718824
Germany                    3.574265
Netherlands                2.144559
Ireland                    2.144559
Australia                  1.667990
France                     1.032566
India                      0.794281
New Zealand                0.635425
Sweden                     0.555997
Switzerland                0.555997
Poland                     0.555997
Italy                      0.555997
Belgium                    0.476569
South Africa               0.476569
Brazil                     0.476569
Israel                     0.397141
Bulgaria                   0.317712
Singapore                  0.317712
Austria                    0.238284
Mexico                     0.238284
Russia                     0.238284
Finland                    0.238284
Greece                     0.158856
Denmark                    0.158856
Portugal                   0.158856
Colombia                   0.158856
Croatia                    0.158856
Moldova                    0.079428
Japan                      0.079428
Romania                    0.079428
Costa Rica                 0.079428
Philippines                0.079428
Bosnia and Herzegovina     0.079428
Czech Republic             0.079428
Zimbabwe                   0.079428
Latvia                     0.079428
Spain                      0.079428
Bahamas, The               0.079428
Uruguay                    0.079428
Norway                     0.079428
Nigeria                    0.079428
Hungary                    0.079428
Thailand                   0.079428
China                      0.079428
Georgia                    0.079428
Slovenia                   0.079428
Name: Country, dtype: float64
```

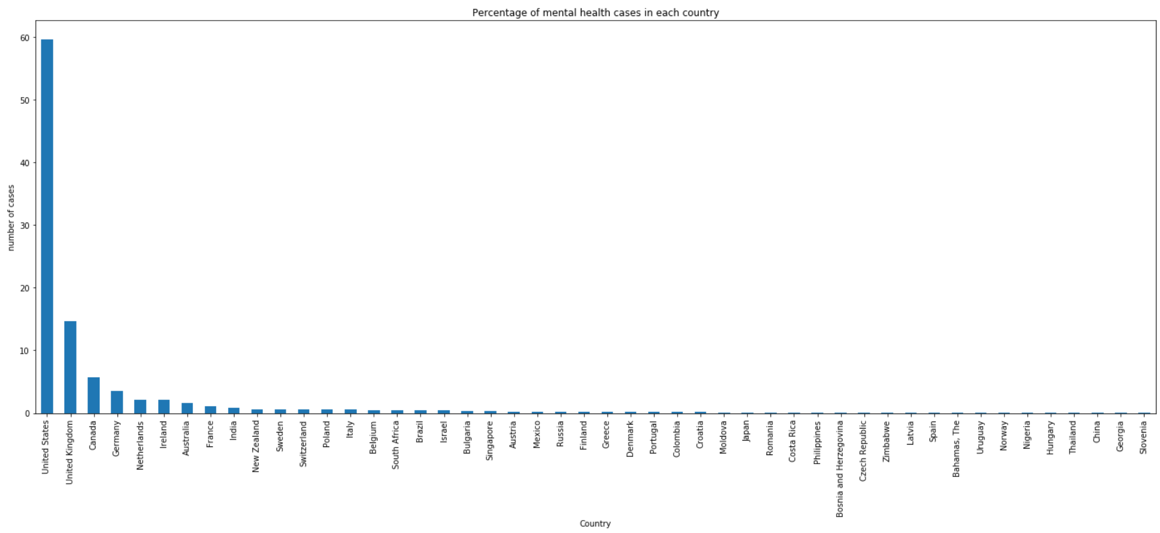
From the above data, we plot the bar graphs to get a better understanding of the numbers and percentages.

1. We plot the percentage of cases in each country.

In [11]:

```
#Plotting x% of cases in each country
f, ax = plt.subplots(figsize=(25,9))
chartTitle = "Percentage of mental health cases in each country"
yaxisLabel = "number of cases"
xaxisLabel = "Country"

cases_by_percent.plot.bar(ax=plt.axes(title=chartTitle,ylabel = yaxisLabel ,xlab
el = xaxisLabel) )
plt.show()
```

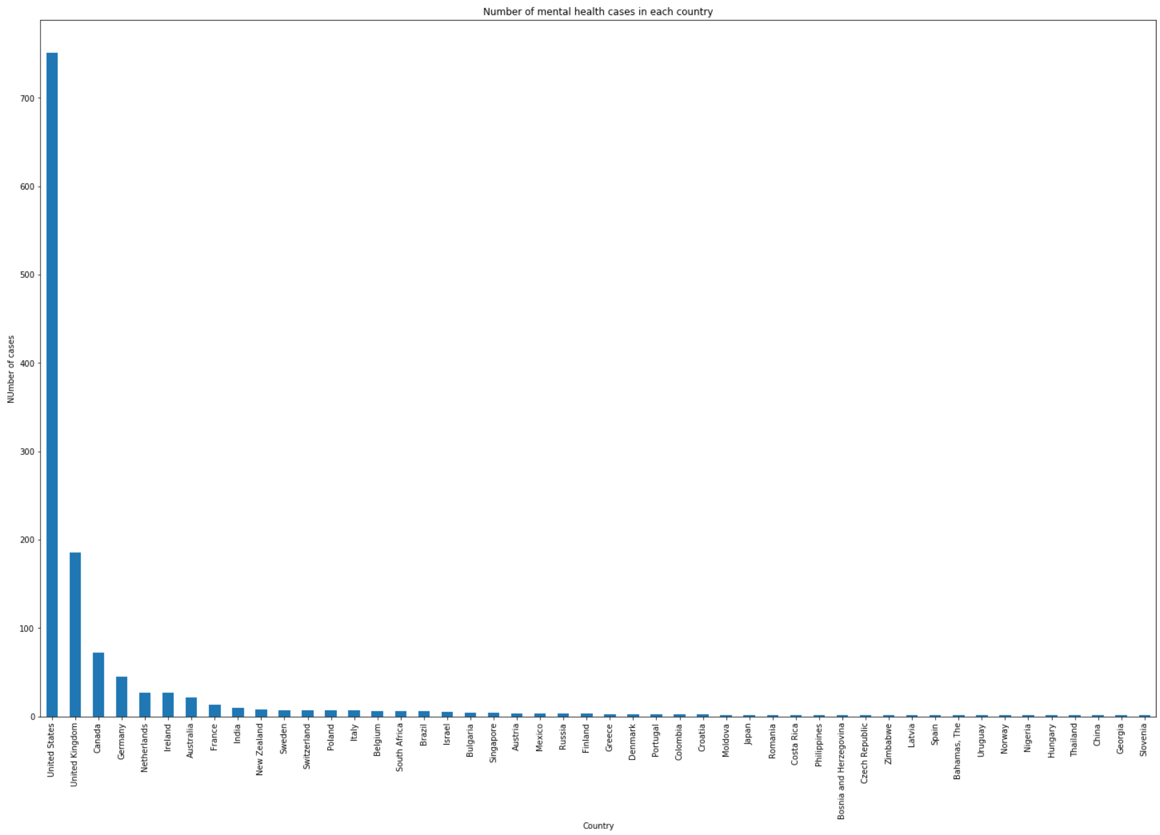


- From the above graph, we get that 59.69% of the mental health cases are in the United States. They hold a significant percentage, followed by United Kingdom with just 14.69%.
- It can be seen that the percentage of cases in Canada, Germany, Netherlands, Ireland, Australia and France are in single digits.
- The cases in other countries are almost negligible in comparison to the above mentioned countries.

1. We plot the cases in numbers recorded in each country.

In [35]:

```
#Plotting cases in each country
f, ax = plt.subplots(figsize=(25,16))
plt.title('Number of mental health cases in each country', fontsize=80)
plt.ylabel('Number of cases', fontsize=70)
plt.ylabel('Country', fontsize=70)
cases.plot.bar(ax=plt.axes(title=chartTitle,ylabel = yaxisLabel ,xlabel = xaxisL
abel) )
plt.show()
```



> As seen in the second bar chart, it is difficult for us to view the values of the values of the countries at the back of the list. Also, after a point the number of cases are 3 or less than that. Stagnant values in the data serve no purpose. So, for this specific purpose, we restrict our analysis for the frequency of cases to the top 20 countries where there are high number of people suffering from the illness.

Hence, we extract the countries in the list which have significant cases. Below, we extract the top 20 countries.

In [13]:

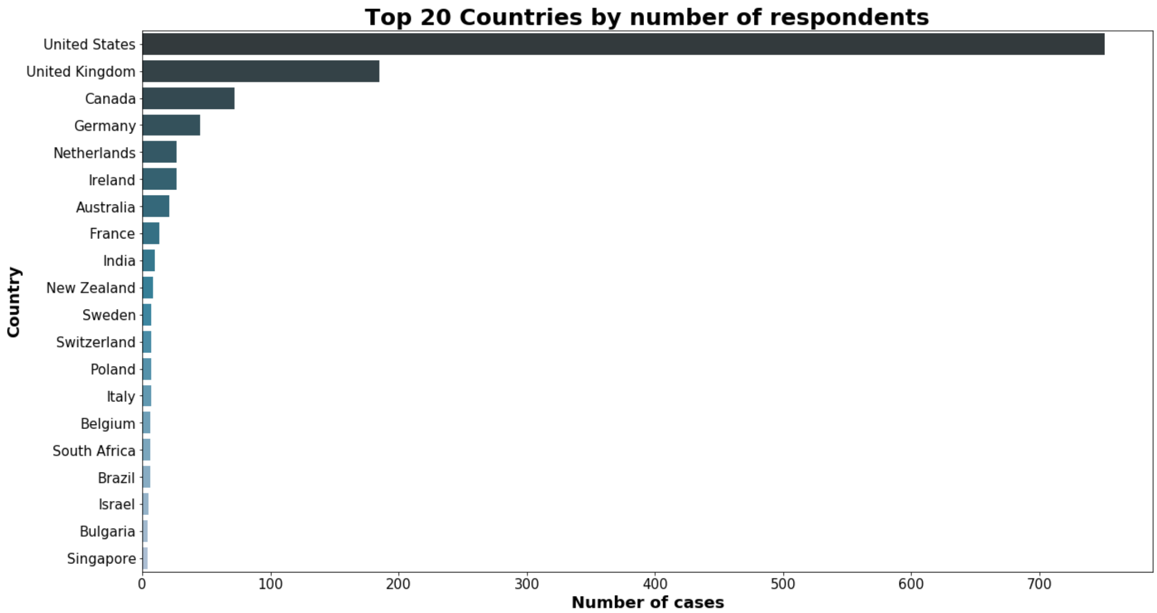
```
# top 20 countries in numbers
top_20_countries = data['Country'].value_counts()[:20]
print(top_20_countries)
```

```
United States     751
United Kingdom    185
Canada             72
Germany            45
Netherlands        27
Ireland            27
Australia          21
France             13
India              10
New Zealand         8
Sweden              7
Switzerland         7
Poland              7
Italy               7
Belgium             6
South Africa        6
Brazil              6
Israel              5
Bulgaria            4
Singapore           4
Name: Country, dtype: int64
```

We plot the result in the form a bar plot as it would give the accurate visualization.

In [14]:

```
#Plot the top 20 countries in the list as a bar graph
top_20_countries = data['Country'].value_counts()[:20].to_frame()
plt.figure(figsize=(20,11))
sns.barplot(top_20_countries['Country'],top_20_countries.index,palette="PuBuGn_d")
plt.title('Top 20 Countries by number of respondents',fontsize=25,fontweight="bold")
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)
plt.ylabel('Country', fontsize=18,fontweight="bold")
plt.xlabel('Number of cases', fontsize=18,fontweight="bold")
plt.show()
```



From the above visualization,

- Out of the total of 1259 cases golbally, 751 cases are in United States, which constitutes about 59.65%.
- It is followed by the United Kingdom with a share of 14.69% (185 cases).
- As can be infered, there is a significantly higher number of cases in the tech workspace in the united states compared to any other country.

ANALYSIS OF THE DATA FOR 'ATTITUDE TOWARDS MENTAL HEALTH ILLNESS BY GEOAGRAPHIC LOCATION'

We live in a society where people consider mental health issues as a stigma. This creates insecurity among individuals suffering from it. This affects their perception towards the illness.

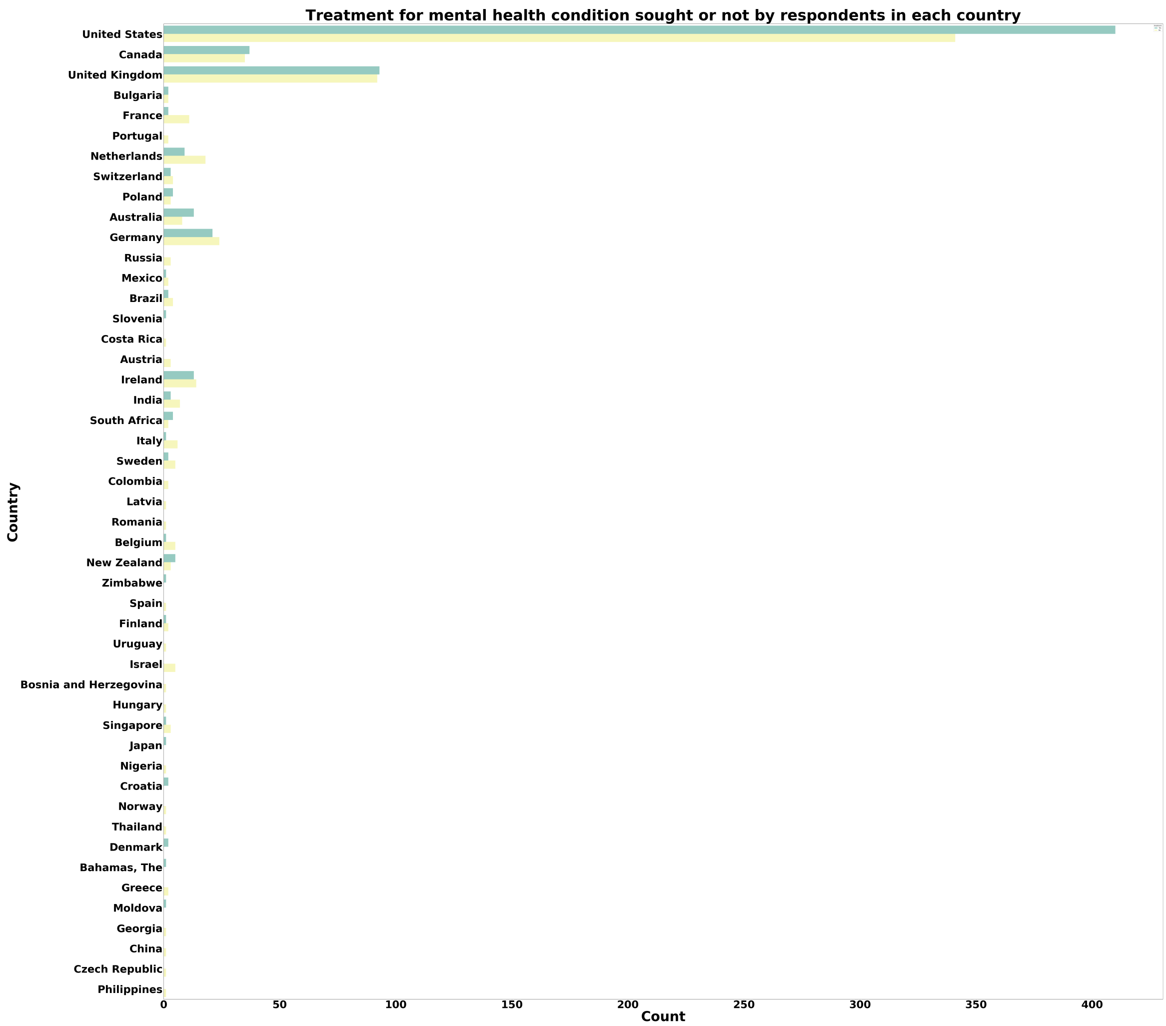
We first plot the graoh between the "treatment" and the "country" columns of the dataset. The idea behind this is that whether a respondent is seeking treatment for his illness or not shows a lot about the his attitude and that of the poeple he/she is surrounded with. We have plotted it w.r.t each country to understand the attitude variation from one country to another.

In [15]:

```
# Plot the graph to understand
plt.figure(figsize=(120,120))
sns.countplot(y = data['Country'], hue = data['treatment'], palette="Set3")
plt.title('Treatment for mental health condition sought or not by respondents in each country', fontsize = 100, fontweight="bold")
plt.xticks(fontsize=70, fontweight="bold")
plt.yticks(fontsize=70, fontweight="bold")
plt.ylabel('Country', fontsize = 90, fontweight="bold")
plt.xlabel('Count', fontsize = 90, fontweight="bold")
```

Out[15]:

```
Text(0.5, 0, 'Count')
```



In the above visualization, the green bar represents a "yes" and a yellow bar represents a "no" for whether a treatment is taken or not by the respondent for his/her mental health condition. From the graph, we can conclude the following:

- The attitudes of people in the United States is very positive. The number of people seeking treatment for their health condition is slightly higher than the ones who don't.
- Whereelse, the number of people who have sought treatment are relatively same in Canada and the United Kingdom.
- However, in Netehrlands and Germany, more respondents have answered "no" for this question related to treatment. For the former one, the difference is slight, where else, for the latter, it is insignificant.
- Same can be observed for India, where the yellow graph is longer than the green graph indicating more no's.

This, hence, represents the attitude of people towards mental health illness across various geographical locations.

ANALYSIS OF THE DATA FOR 'STRONGEST PREDICTORS OF MENTAL HEALTH ILLNESS IN WORKPLACE'

The aim of reseraching into investigating the predictors of mental health illness at workplaces is to achieve a declination in the number of cases. Analysing the data to understand the attitudes of people towards mental disorder at workplaces is also significant. This helps in improving help-seeking services and support facilities. Also, the attitudes of people towards this condition plays a vital role in the recovering of the patient suffering from it.

Predictor 1

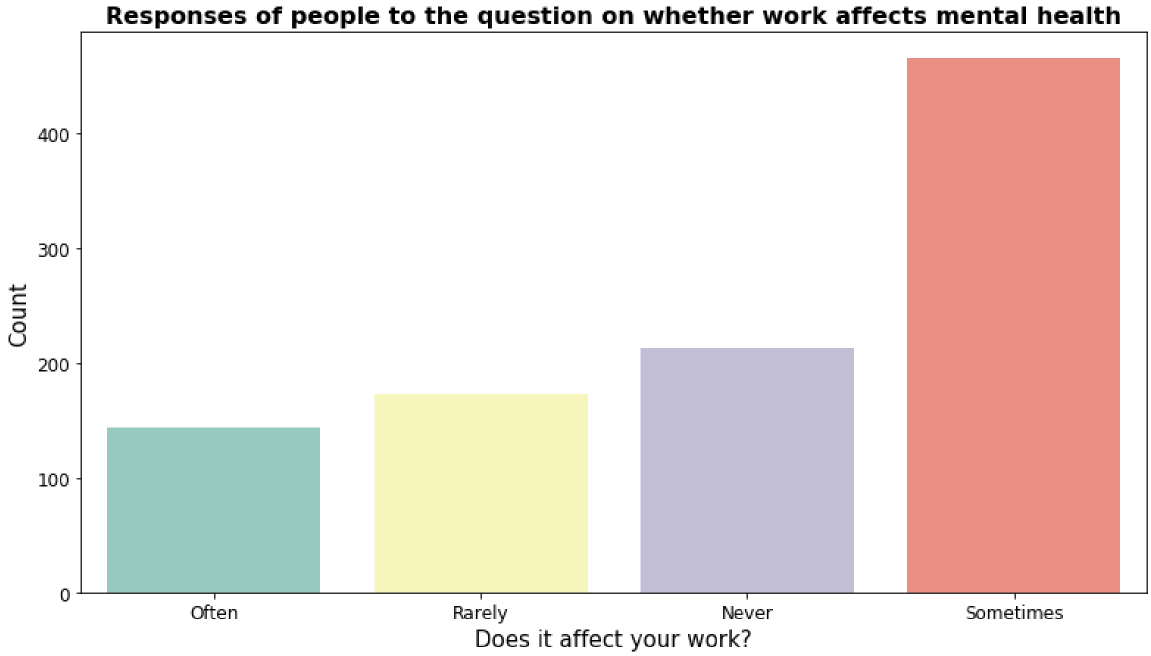
One way of predicting the mental health illness at an early stage can be by keeping a track on the work efficiency of the employees. To solve the purpose, we plot the variation in the responses to the question "Does mental health illness affect the work?".

In [16]:

```
#calculating the response to the question "does MHI affect the work"
fig,ax  =plt.subplots(figsize=(13,7))
sns.countplot(data['work_interfere'], palette="Set3")
plt.title('Responses of people to the question on whether work affects mental he
alth', fontsize=16,fontweight="bold")
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.ylabel('Count', fontsize=15)
plt.xlabel('Does it affect your work?', fontsize=15)
```

Out[16]:

```
Text(0.5, 0, 'Does it affect your work?')
```



From the above visualization, we can conclude that mental health problems affects the working efficiency of an individual. As in the above graph, we can see that the highest number of survey responders said "Sometimes". And so, employers can predict it to some extent.

Predictor 2

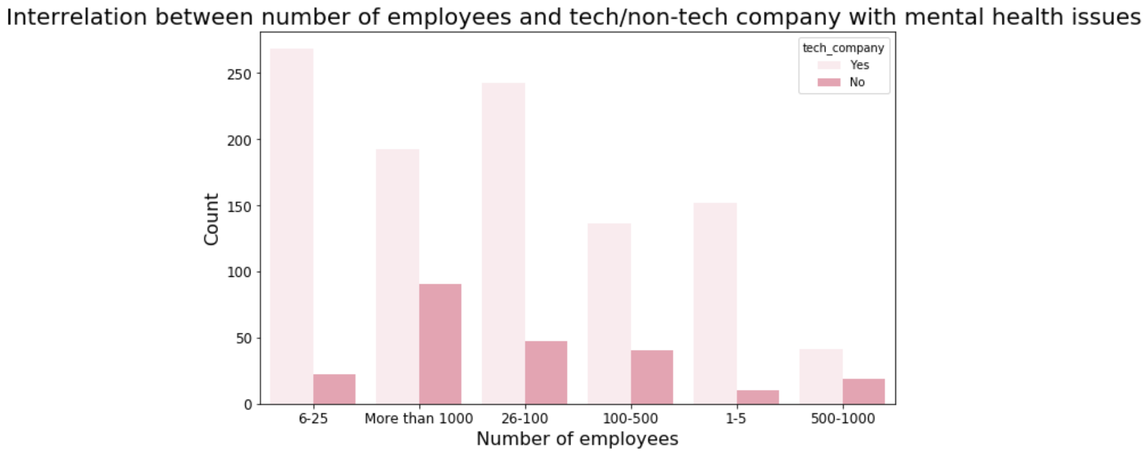
Another predictor can be the type of workplace environment, understanding that whether the workplace is a tech company or not and the number of employees working in that organization. We try to understand whether these factors affect an individuals mental status. Here, we will check the frequency of mental health illness by type of workplace and number of employees in terms of they are in a tech company or not.

In [60]:

```
# Understanding the relation between "tech_company", "no_employees" and number o
f cases
fig,ax  =plt.subplots(figsize=(10,6))
sns.countplot(x = data['no_employees'], hue = data['tech_company'],color='#EE99A
C')
plt.title('Interrelation between number of employees and tech/non-tech company w
ith mental health issues ', fontsize = 20)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.xlabel('Number of employees', fontsize = 16)
plt.ylabel('Count', fontsize = 16)
```

Out[60]:

```
Text(0, 0.5, 'Count')
```



- The above graph shows that the odds of a person suffering from a mental health issue in a tech company is very high in comparison to the ones in a non-tech company, irrespective of the strength of the working staff of the organization.
- The lowest number of cases are observed when there are no more than 5 employees in a non-tech workplace.
- The highest number of cases can be seen when the employee number varies between 6-25 and works in a tech company. However, with the same number of employees, but in a non-tech company, the number of cases observed are drastically less.

ANALYSIS OF THE DATA FOR 'ATTITUDE TOWARDS MENTAL HEALTH ILLNESS IN WORKPLACE'

Attitude of the company towards mental health illness can be identified from the benefits and ease of work, job security, recovering time it provides to its employees suffering from mental disorders. The perception and the behaviour of the coworkers at workplace is also signficant. Therefore, to address the question

Variable 1

We use the variable "benefits" to know whether the employer provides benfits to the mental health sufferers or not.

In [44]:

```
#calculating answers for each of the three responses for the question "benefits"
data['benefits'].value_counts()
```

Out[44]:

```
Yes           477
Don't know    408
No            374
Name: benefits, dtype: int64
```

In [45]:

```
#calculating answers for each of the three responses for the question "benefits"
data['benefits'].value_counts(normalize=True)
```

Out[45]:

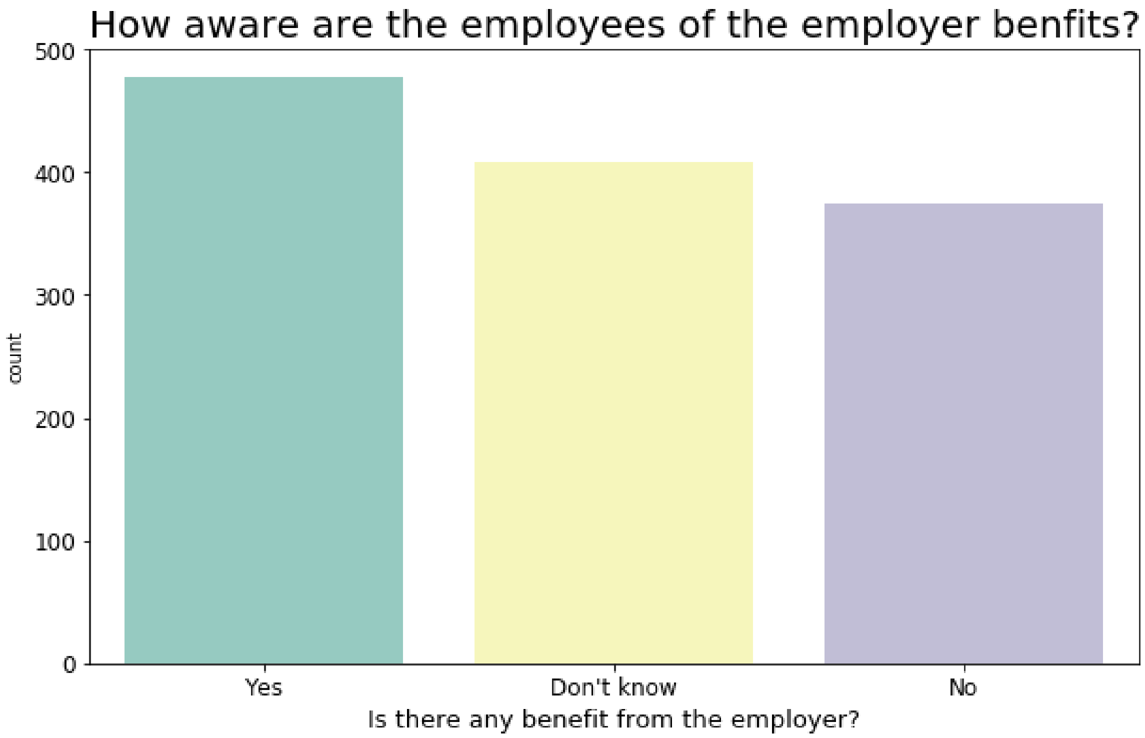
```
Yes           0.378872
Don't know    0.324067
No            0.297061
Name: benefits, dtype: float64
```

In [61]:

```
# Plotting the benefits awareness graph
fig,ax  =plt.subplots(figsize=(10,6))
sns.countplot(data['benefits'], palette='Set3')
plt.title('How aware are the employees of the employer benfits?', fontsize = 20)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.xlabel('Is there any benefit from the employer?', fontsize = 13)
```

Out[61]:

```
Text(0.5, 0, 'Is there any benefit from the employer?')
```



- From the above visualization, we can get that 477 people answer "Yes", which is considered to be good.
- We also get that approximately 400 employees are unaware whether their employers assist them with any mental health related benefits or not. This shows the attitude of the employers and their ignorant nature and that they should do more to spread awareness of the welfare they provide.
- From the code above, we also get to know that approximately 29% of the respondents answer "No" which implied the attitude of the employers towards the illness and the fact that it is not taken seriously.

Variable 2

Another variable to understand the attitude is "wellness_program", whether the employer has ever discussed mental health as part of a wellness program.

In [41]:

```
data['wellness_program'].value_counts()
```

Out[41]:

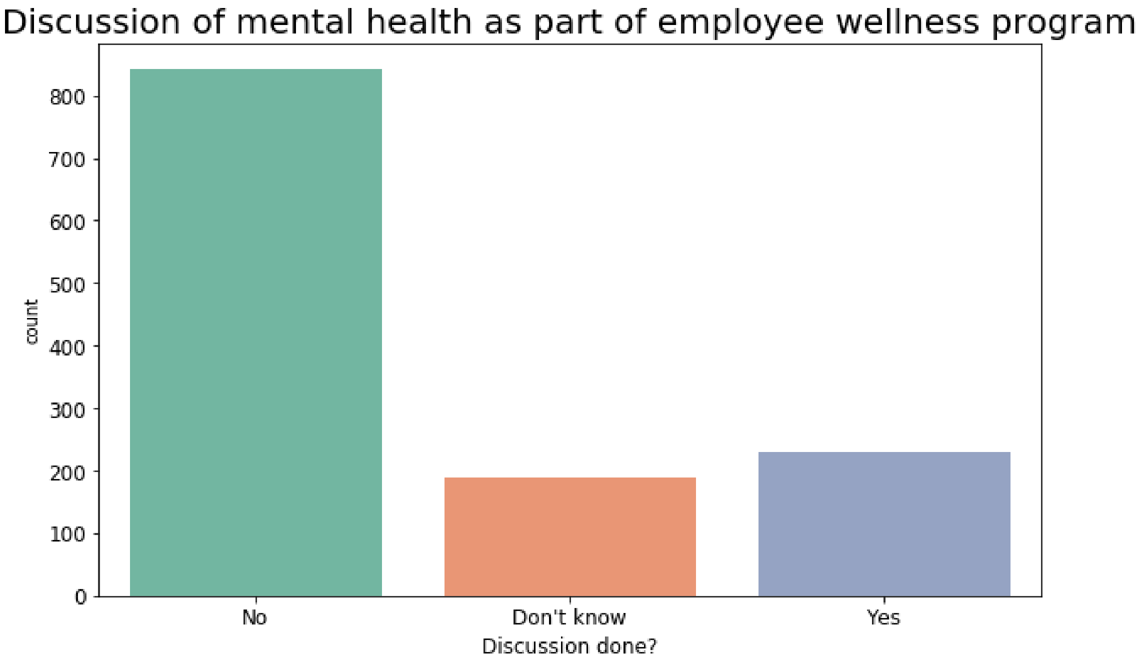
```
No              842
Yes             229
Don't know      188
Name: wellness_program, dtype: int64
```

In [62]:

```
fig,ax  =plt.subplots(figsize=(10,6))
sns.countplot(data['wellness_program'], palette='Set2')
plt.title('Discussion of mental health as part of employee wellness program', fo
ntsize = 20)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.xlabel('Discussion done?', fontsize = 12)
```

Out[62]:

```
Text(0.5, 0, 'Discussion done?')
```



From the above visulization it can be observed that out of a total of 1259 respondents, 842 had answered "No". This is a matter of concern and needs to be addressed. Employers, on their part need to spread awareneess.

Variable 3

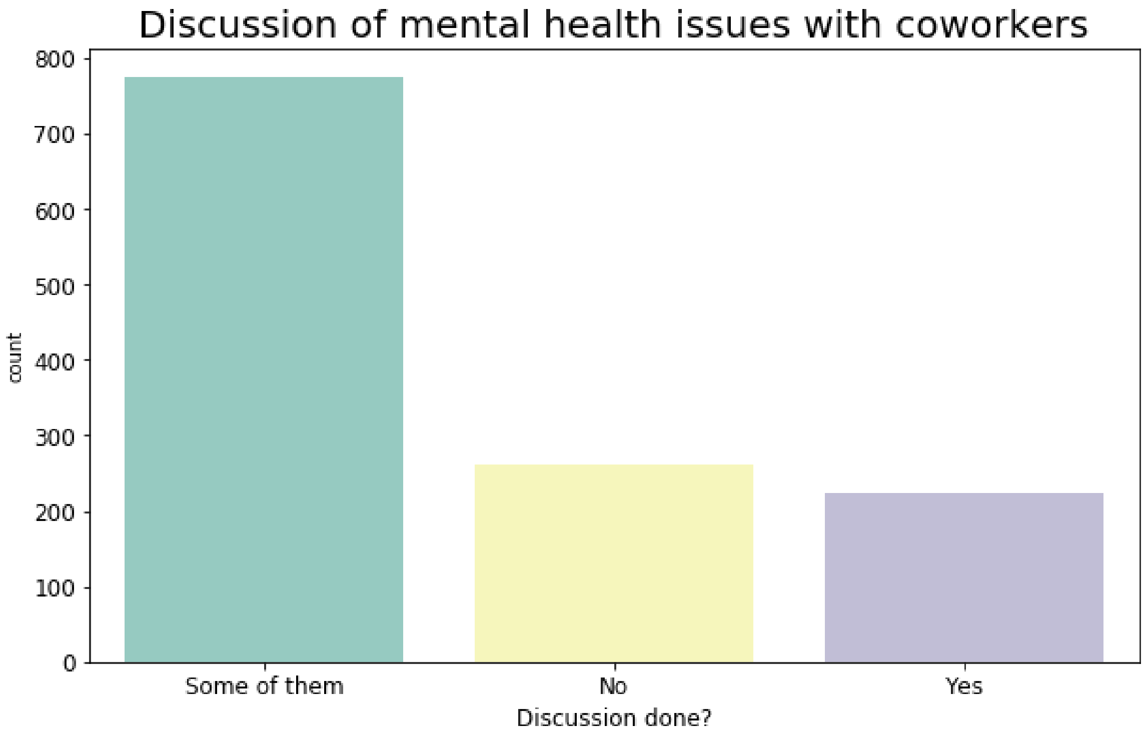
A variable that helps depict the attitude of coworkers is "coworkers". It includes how open are the respondents to talk about mental health issues with their coworkers.

In [63]:

```
# Plot the graph to understand how many respondents would discuss mental health
 issues with coworkers
fig,ax  =plt.subplots(figsize=(10,6))
sns.countplot(data['coworkers'], palette='Set3')
plt.title('Discussion of mental health issues with coworkers', fontsize = 20)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.xlabel('Discussion done?', fontsize = 12)
```

Out[63]:

```
Text(0.5, 0, 'Discussion done?')
```



The above graph shows that more than half the number of people are willing to discuss about mental health issues with their coworkers. It shows that the attitude of the coworkers towards this illness is positive and makes the respondent feel secure enough to talk about it.

Variable 4

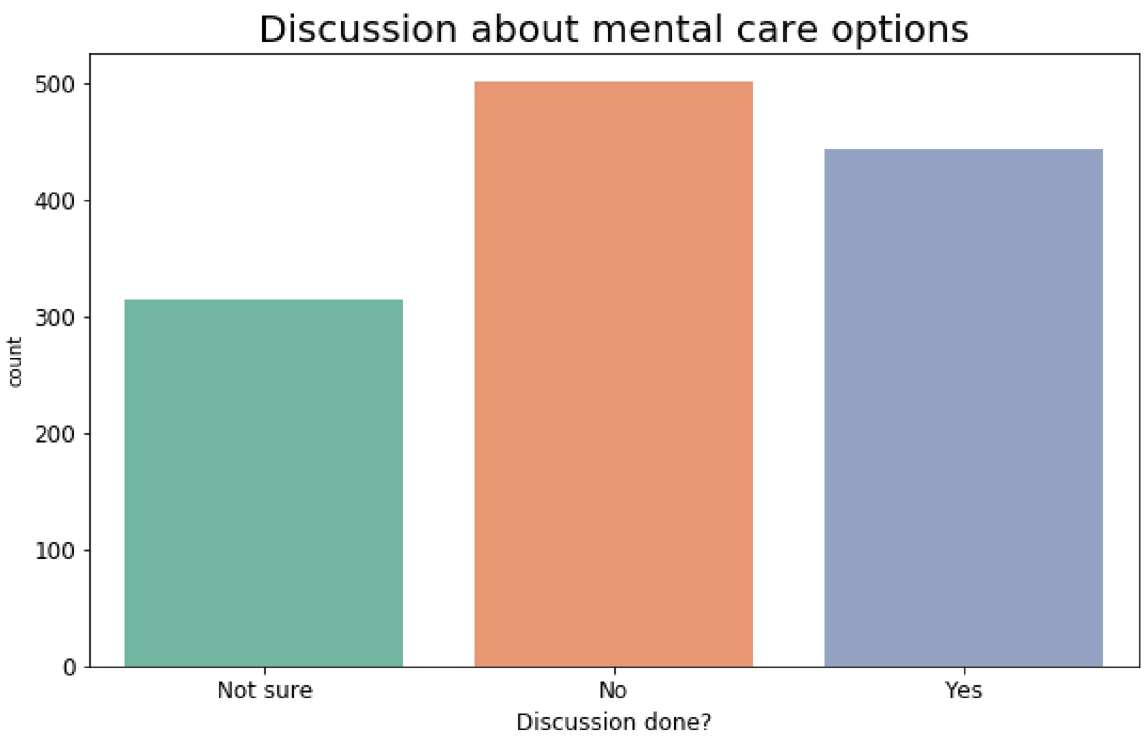
Whether the respondent is aware about the mental care options provided by the employer or not says a lot about the atttude of the respondent himself and that of the employer.

In [72]:

```
# Plot and count "care_options" awareness
fig,ax  =plt.subplots(figsize=(10,6))
sns.countplot(data['care_options'], palette='Set2')
plt.title('Discussion about mental care options', fontsize = 20)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.xlabel('Discussion done?', fontsize = 12)
```

Out[72]:

```
Text(0.5, 0, 'Discussion done?')
```



It can be inferred that about 500 respondents do not discuss about the mental care options provided by the employer. This shows ignorant attitude of the employer.

Variable 5

"Would you bring up a mental health issue with a potential employer in an interview?". Plotting this question helps us understand the attitude of people at the workplace towards mental health illness, and how safe do employees or interviewees feel to talk about it and accept it.

In [58]:

```
data['mental_health_interview'].value_counts(normalize=True)
```

Out[58]:

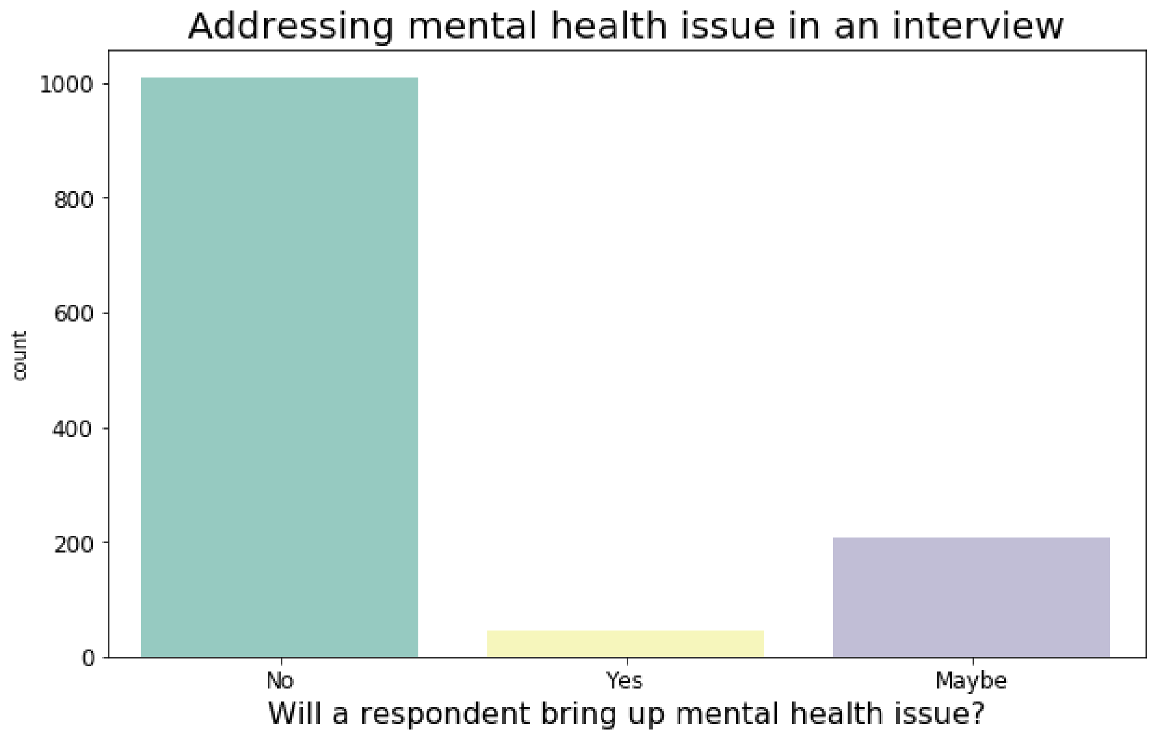
```
No       0.800635
Maybe    0.164416
Yes      0.034948
Name: mental_health_interview, dtype: float64
```

In [65]:

```
# Plot how likely is a respondent to talk about mental health in an interview
fig,ax  =plt.subplots(figsize=(10,6))
sns.countplot(data['mental_health_interview'], palette="Set3")
plt.title('Addressing mental health issue in an interview', fontsize = 20)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.xlabel('Will a respondent bring up mental health issue?', fontsize = 16)
```

Out[65]:

```
Text(0.5, 0, 'Will a respondent bring up mental health issue?')
```



From the above visualization, we get that 80% of the respondents are likely to not bring up the issue of mental health in an interview. This shows how "uncomfortable" it is considered.

Variable 6

"Would you bring up a physical health issue with a potential employer in an interview?". This question is asked to compare the difference between the responses when it comes to physical and mental health issue.

In [74]:

```
data['phys_health_interview'].value_counts(normalize=True)
```

Out[74]:

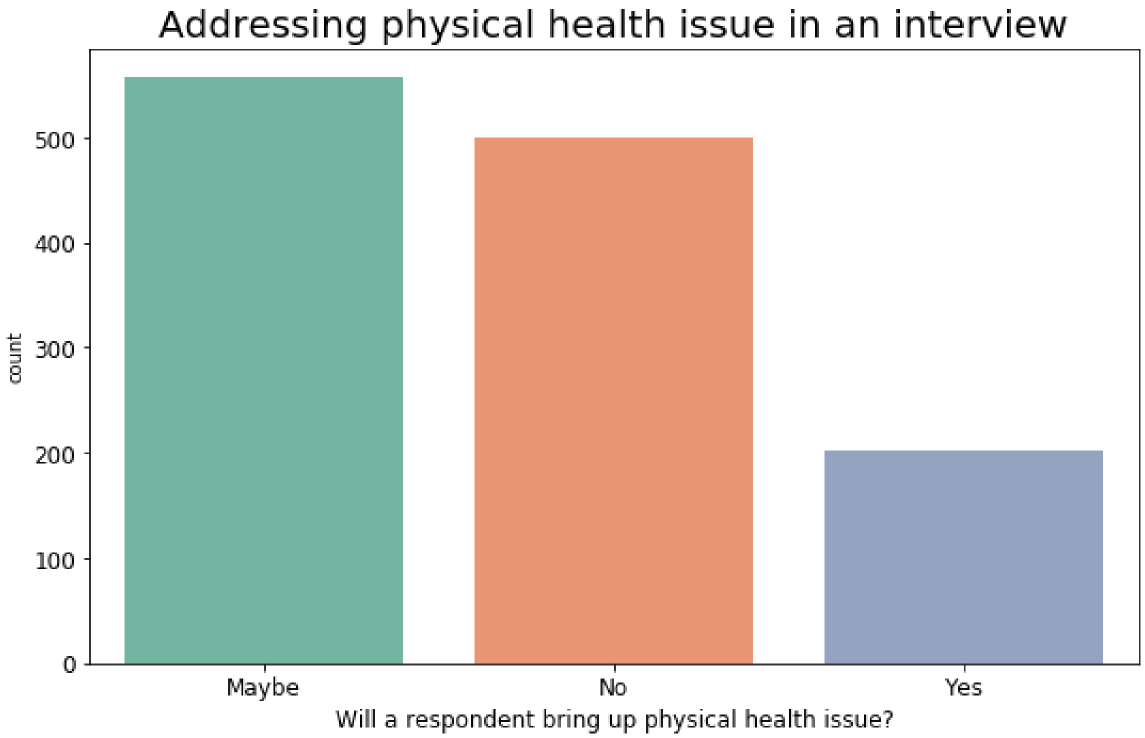
```
Maybe    0.442415
No       0.397141
Yes      0.160445
Name: phys_health_interview, dtype: float64
```

In [73]:

```
# Plot how likely is a respondent to talk about physical health issue in an inte
rview
fig,ax  =plt.subplots(figsize=(10,6))
sns.countplot(data['phys_health_interview'], palette="Set2")
plt.title('Addressing physical health issue in an interview', fontsize = 20)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.xlabel('Will a respondent bring up physical health issue?', fontsize = 12)
```

Out[73]:

```
Text(0.5, 0, 'Will a respondent bring up physical health issue?')
```



On comparing the "yes's" of both the above questions in the survey on whether the respondent will bring up the issue of mental health and physical health with a potential employer or not, we observe taht for the former, it is just 3% where else for the latter, it is 16%. The difference is very high. It depicts the attitude towards mental health.

Variable 7

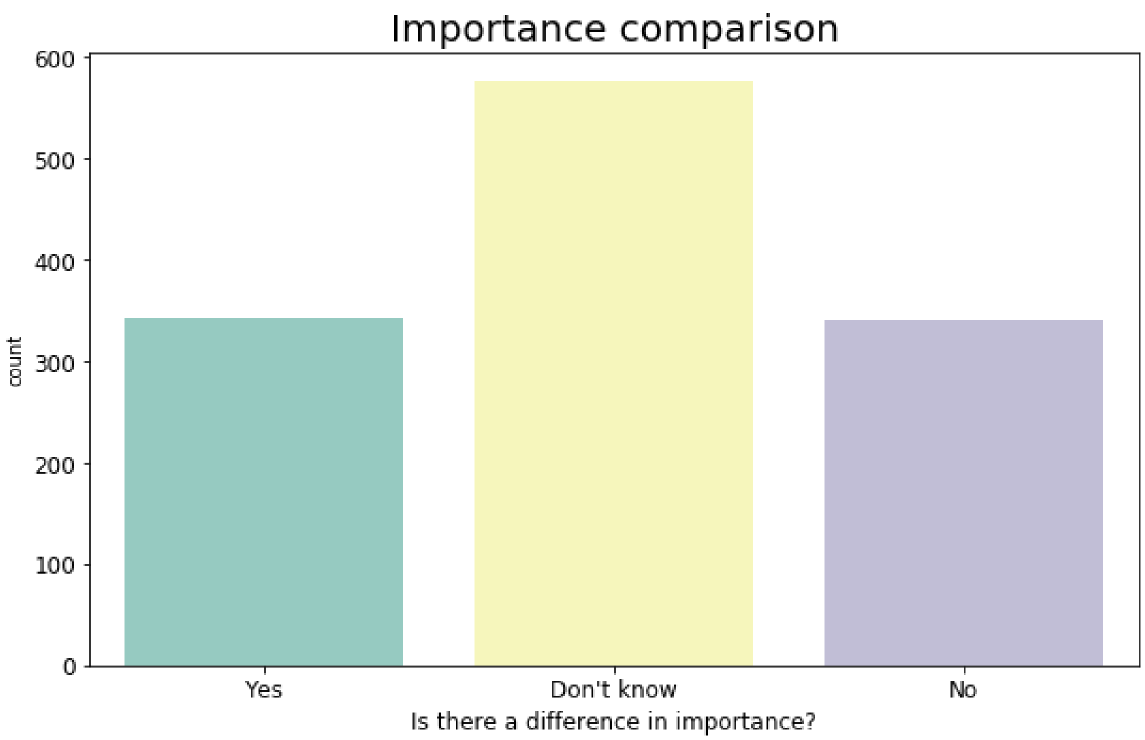
The "mental_vs_physical" variable talks about the difference in the importance given by the employer to mental and physical health issues, as believed by the respondents.

In [78]:

```
# Plot "mental_vs_physical"
fig,ax  =plt.subplots(figsize=(10,6))
sns.countplot(data['mental_vs_physical'], palette="Set3")
plt.title('Importance comparison', fontsize = 20)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.xlabel('Is there a difference in importance?', fontsize = 12)
```

Out[78]:

```
Text(0.5, 0, 'Is there a difference in importance?')
```



> The above two graphs showing whether a respondent will talk about a mental health issue or a physical health issue with a potential employer or not , and the difference in the number of "yes's" shows the discrimination. However, this does not give us some harcore conclusion.

INSIGHTS

The insights from the above analysis are as follows:

- Out of the total 1259 respondents of the survey, 751 cases are observed in the United States constituting about 59.65%, followed by United Kingdom with 14.69% (185 cases).
- Attitudes towards mental health illness varies by geographical location. Number of employees seeking treatment are slightly higher than the ones who don't in the United States. Whereas, the numbers are same in Canada and United Kingdom. However, in Netherlands and Germany, more respondents have answered "no" for this question related to treatment. For the former, the difference is minute, whereas, for the latter, it is insignificant. Same can be observed for India, where the yellow graph is longer than the green graph indicating more no's . Hence, governments and companies should focus on awareness development programs and in decreasing the stigma it invokes and encourage people to take medical assistance.
- Poor mental health impacts an individual's workplace productivity. Maintaining records of employee performance can help in the early identification of the problem, if any.
- Frequency of cases in tech companies are very high in comparison to that of non-tech companies irrespective of the workforce of the organisation. This clearly indicates that employers of tech companies need to tackle the root causes to decrease these numbers.
- 32.4% of the employees are unsure about the employer benefits and 29.7% are completely unaware about it. Employers need to increase the employee knowledge about the available mental health support services they are entitled to.
- 66.87% of the workforce said that mental health wasn't discussed in the employee wellness programs. Employers need to open up conversations around mental health and well-being and the health care options available.
- Many employees would proactively approach some of their co-workers to discuss their mental health illness. However, employers can work more into creating a culture where the employees can openly talk about it with any of their colleagues.
- Employees do not feel comfortable in bringing up their mental issues in an interview with an employer. The stigma that surrounds this issue needs to be curbed to allow open discussions and revelations.
- It is the employer's responsibility to create a culture of awareness and support for mental health illness in workspaces.

In [ ]:

```

```

QUESTION 2. What were the top Australian news topics over the last decade, and what can these say about the national conversation?