In [1]:

```
# !pip install Pyppeteer
# !pyppeteer-install
```

In [2]:

```
import plotly.io as pio
#pio.renders.default = 'notebook'
```

Analysis of Gender Inequality across World

Data is sourced from World Data Bank, Census, and US Bureau of Labor Force Statistics. The data was narrowed down to include countries depending on their development indicators.

The least developed nations - Yemen and Afghanistan; developing nations - India and Azerbaijan; developed nations - United States.

Import all the necessary libraries

In [3]:

```
import pandas as pd
import altair as alt
from IPython.display import HTML
import matplotlib.pyplot as plt
import geopandas
```

In [4]:

```
alt.data_transformers.enable('default', max_rows=None)
```

Out[4]:

```
DataTransformerRegistry.enable('default')
```

Load all Datasets

In [5]:

```
jobs = pd.read_csv('JobsData.csv')
parliament = pd.read_csv('Par_Women_Data.csv')
women_wage_perc = pd.read_excel('wage_per_occupation.xlsx', sheet_name="Table 14")
lp = pd.read_csv("Labor Force Participation Rate of Mothers and Fathers by Age of Younge
                          skiprows=1)
world_data = pd.read_csv("WDIData.csv")
mortality  = pd.read_csv("MaternalMortalityData.csv")
inequality  = pd.read_csv("gender-inequality-index-from-the-human-development-report.csv
```

Data Preprocessing

In [6]:

```
jobs = jobs.rename(columns = {"Indicator Name":"Variables"})
```

In [7]:

```
jobs.head(3)
```

Out[7]:

	Country Name	Country Code	Variables	Indicator Code	1990	1991	1992	1993	1994	199
0	Arab World	ARB	Access to electricity (% of population)	EG.ELC.ACCS.ZS	74.384239	74.382220	74.313160	75.349325	75.788522	76.21413
1	Arab World	ARB	Adolescent fertility rate	SP.ADO.TFRT	69.467160	68.211985	67.314595	65.256059	63.177552	60.90790

			(births per 1,000 wo...							
2	Arab World	ARB	Age dependency ratio (% of working-age populat...	SP.POP.DPND	87.481340	86.726178	86.058118	84.906750	83.598142	81.94641

3 rows × 31 columns

Dropping unnecessary columns and extracting only the percentage of Male and female employment in three sectors: \

```
1. Agriculture \
2. Industry \
3. Services
```

```
In [8]: job_list_of_values = ["Employment in agriculture (% of total employment) (modeled ILO es
                             "Employment in agriculture, female (% of female employment) (modeled I
                             "Employment in agriculture, male (% of male employment) (modeled ILO e
                             "Employment in industry (% of total employment) (modeled ILO estimate)
                             "Employment in industry, female (% of female employment) (modeled ILO 
                             "Employment in industry, male (% of male employment) (modeled ILO esti
                             "Employment in services (% of total employment) (modeled ILO estimate)
                             "Employment in services, female (% of female employment) (modeled ILO 
                             "Employment in services, male (% of male employment) (modeled ILO esti
                             "Labor force with advanced education, female (% of female working-age 
                             "Labor force with basic education, female (% of female working-age pop
                             "Labor force with intermediate education, female (% of female working-
                             "Labor force participation rate, female (% of female population ages 1
                             "Fertility rate, total (births per woman)",
                      "Literacy rate, adult female (% of females ages 15 and above)",
                      "Literacy rate, adult male (% of males ages 15 and above)",
                      "Self-employed, female (% of female employment) (modeled ILO estimate)",
                      "Self-employed, male (% of male employment) (modeled ILO estimate)",
                       ]
        jobs_df = jobs[jobs['Variables'].isin(job_list_of_values)]
```

```
In [9]: jobs_df_small = jobs_df.reset_index()
        jobs_df_small = jobs_df_small.drop(columns = ['Indicator Code','index'])
        jobs_dfp = jobs_df_small.pivot(index='Variables', columns=['Country Name', 'Country Code
```

```
In [10]: jDF = jobs_dfp
         jDF = jobs_dfp.rename(columns={"Employment in agriculture (% of total employment) (model
                             "Employment in agriculture, female (% of female employment) (modeled I
                             "Employment in agriculture, male (% of male employment) (modeled ILO e
                             "Employment in industry (% of total employment) (modeled ILO estimate)
                             "Employment in industry, female (% of female employment) (modeled ILO 
                             "Employment in industry, male (% of male employment) (modeled ILO esti
                             "Employment in services (% of total employment) (modeled ILO estimate)
                             "Employment in services, female (% of female employment) (modeled ILO 
                             "Employment in services, male (% of male employment) (modeled ILO esti

                             "Labor force with advanced education, female (% of female working-age 
                             "Labor force with basic education, female (% of female working-age pop
                             "Labor force with intermediate education, female (% of female working-
                             "Labor force participation rate, female (% of female population ages 1

                       "Fertility rate, total (births per woman)":'Fertility',
                      "Literacy rate, adult female (% of females ages 15 and above)":'lit_F',
                      "Literacy rate, adult male (% of males ages 15 and above)":'lit_m',
                      "Self-employed, female (% of female employment) (modeled ILO estimate)":'self
                      "Self-employed, male (% of male employment) (modeled ILO estimate)":'self_Emp
```

In [11]:

```
jDF.reset_index(inplace=True)
```

In [12]:

```
jDF.head()
```

Out[12]:

Variables	level_0	Country Name	Country Code	Agriculture_Total	Agriculture_Female	Agriculture_Male	Industry_Total	Ind
0	1990	Arab World	ARB	NaN	NaN	NaN	NaN	
1	1990	East Asia & Pacific	EAS	NaN	NaN	NaN	NaN	
2	1990	East Asia & Pacific (excluding high income)	EAP	NaN	NaN	NaN	NaN	
3	1990	Euro area	EMU	NaN	NaN	NaN	NaN	
4	1990	Europe & Central Asia	ECS	NaN	NaN	NaN	NaN	

5 rows × 21 columns

In [13]:

```
#renaming columns with appropriate names

jDF = jDF.rename(columns={'level_0':'Year',
                          "Country Name":"Country",
                          "Country Code":"CODE"})
```

In [14]:

```
#creating a list of year values

years = jDF['Year'].unique() # get unique field values
years = list(filter(lambda x:  x > '2000', years)) # filter out None values
years.sort() # sort alphabetically
```

In [15]:

```
#binding values to drop-down
input_dropdown = alt.binding_select(options=years)

selectYear = alt.selection_point(
    name='Select',
    fields=['Year'],
    value='2016',
    bind=input_dropdown
    #bind=alt.binding_range(min=1990, max=2016)
)
```

In [16]:

```
display(HTML("""
<style>
form.vega-bindings {
  position: absolute;
  left: 0px;
  top: 0px;
}
</style>
"""))
```

In [17]:

```
#renaming legend names appropriately


legend_labels = ("datum.label == 'Agriculture_Female' ? 'Agriculture' : datum.label == '
axis_labels = ("datum.label == 'Agriculture_Female' ? 'Female' : datum.label == 'Industr

#selection of color palette

color_category =['#3A2A51','#52A675','#FF595E'] #3 distinct
color_category1_light = ['#3A2A51','#BFAED5'] #2 lighter shade of 1 category color
color_category2_light = ['#52A675','#9FD0B4']
color_category3_light = ['#FF595E','#FFADB0']
heatmap = ['#3A2A51', '#FFC2C4']
heatmap1 = ['#FFC2C4','#3A2A51']
color_two_category = ['#3A2A51','#FF595E'] #2 distinct
#['#6A4C93','#1982C4','#FF924C']
#['#FF6B6B','#4ECDC4','#1A535C']#, '#638ccc'] #distinct; category
#['#000075','#f58231','#800000']
```

What is the share of women employment by sectors?

In [18]:

```
#choosing a stac bar visual

stackedbar = alt.Chart(jDF).mark_bar().add_params(selectYear).transform_filter(selectYea
).transform_fold(
    ['Agriculture_Female','Industry_Female','Service_Female']
).transform_filter(alt.FieldOneOfPredicate(field='Country',
                                           oneOf=['India','Azerbaijan','United States',
                                                  'Afghanistan','Yemen, Rep.']) #'Yemen,
).encode(
    alt.Y('Country:N',
          sort=['Afghanistan','Yemen, Rep.','India','Azerbaijan','United States'], title
    alt.X('value:Q',
          title="Female share(%)", axis=alt.Axis(tickMinStep = 100),
          scale= alt.Scale(domain=[0,100])),
    alt.Color('key:N',
              legend=alt.Legend(orient='right', titleOrient='top',
                                title='Employment Sector',labelExpr=legend_labels),
              scale=alt.Scale(#domain=['Agriculture_Female','Industry_Female','Service_F
                              range= color_category)),
    alt.Order('key:N', sort='ascending'),
    alt.Tooltip('value:Q',format='.1f')
).properties(
    width = 750,
    height = 120,
    title = 'Share of Female Employment in Sectors(%)'
)


text = alt.Chart(jDF).mark_text(color='white',align='center',dx=-14,dy=0,fontSize=11
).transform_filter(
    selectYear
).transform_fold(
    ['Agriculture_Female','Industry_Female','Service_Female']
).transform_filter(alt.FieldOneOfPredicate(field='Country',
                                           oneOf=['India','Azerbaijan','United States',
                                                  'Afghanistan','Yemen, Rep.'])
).encode(
    alt.Y('Country:N',sort=['Afghanistan','Yemen, Rep.','India','Azerbaijan','United Sta
    alt.X('value:Q', stack='zero', scale= alt.Scale(domain=[0,100])),
    alt.Text('value:N',format='.1f'),
    alt.Order('key:N', sort='ascending'),
)


stackedbarsector = alt.layer(
    stackedbar,text
).resolve_scale(
    color='independent'
)
```

In [19]:

```
agri = alt.layer(
  alt.Chart().mark_bar().transform_fold(
    ['Agriculture_Male','Agriculture_Female']
    ).encode(
        alt.Y('key:N',stack='zero',axis=alt.Axis(labelExpr=axis_labels), title = None),
        alt.X('value:Q',
              title = None, axis=None,
           #    axis=alt.Axis(tickMinStep = 100),
               scale=alt.Scale(domain=[0,100])),
        alt.Color('key:N',scale=alt.Scale(range=color_category1_light),legend=None),
      alt.Tooltip('value:Q',format='.1f')

      )
    ,
  alt.Chart().mark_text(color='black',align='center',dx=9.5,dy=0,fontSize=10
    ).transform_fold(
      ['Agriculture_Male','Agriculture_Female']
    ).encode(
        alt.Y('key:N',stack='zero', title = None),
        alt.X('value:Q',stack='zero', title = None),
        alt.Text('value:N',format='.1f')
    )
).properties(
     width = 130,
    height = 50
).facet(
  data=jDF,
     columns=5,
  column =alt.Column('Country:N', title='Male and Female Share in Employment Sectors(%)'
                     header=alt.Header(titleFontSize=15, labelFontSize=12),
                     sort=['Afghanistan','Yemen, Rep.','India','Azerbaijan','United Stat
)

indu = alt.layer(
  alt.Chart().mark_bar().transform_fold(
    ['Industry_Male','Industry_Female']
    ).encode(
        alt.Y('key:N',stack='zero', axis=alt.Axis(labelExpr=axis_labels),title = None),
        alt.X('value:Q',title = None, axis=None,
#               axis=alt.Axis(tickMinStep = 100),
           scale=alt.Scale(domain=[0,100])),
        alt.Color('key:N',scale=alt.Scale(range=color_category2_light),legend=None),
      alt.Tooltip('value:Q',format='.1f')
      )
    ,
  alt.Chart().mark_text(color='black',align='center',dx=12,dy=0,fontSize=10
    ).transform_fold(
    ['Industry_Male','Industry_Female']
    ).encode(
        alt.Y('key:N',stack='zero', title = None),
        alt.X('value:Q',stack='zero', title = None),
        alt.Text('value:N',format='.1f')
    )
).properties(
     width = 130,
    height = 50
```

```
).facet(
  data=jDF,
     columns=5,
  column =alt.Column('Country:N',title=None,header=alt.Header(labels=False),
                     sort=['Afghanistan','Yemen, Rep.','India','Azerbaijan','United Stat
)

serv = alt.layer(
  alt.Chart().mark_bar().transform_fold(
    ['Service_Male','Service_Female']
    ).encode(
        alt.Y('key:N',stack='zero', axis=alt.Axis(labelExpr=axis_labels),title = None),
        alt.X('value:Q',title = None, axis=None,
               #axis=alt.Axis(tickMinStep = 100),
              scale=alt.Scale(domain=[0,100])),
        alt.Color('key:N',scale=alt.Scale(range=color_category3_light),legend=None),
      alt.Tooltip('value:Q',format='.1f')
    )
    ,
  alt.Chart().mark_text(color='black',align='center',dx=-2,dy=0,fontSize=10,
      ).transform_fold(
    ['Service_Male','Service_Female']
    ).encode(
        alt.Y('key:N',stack='zero', title = None),
        alt.X('value:Q',stack='zero'),
        alt.Text('value:N',format='.1f')
    )
).properties(
     width = 130,
    height = 50,
).facet(
  data=jDF,
     columns=5,
  column =alt.Column('Country:N',title=None,header=alt.Header(labels=False),
                     sort=['Afghanistan','Yemen, Rep.','India','Azerbaijan','United Stat

)
```

In [20]:

```
employment_sector = alt.vconcat(stackedbarsector , agri , indu, serv
).resolve_scale(
    color='independent'
).transform_filter(
    alt.FieldOneOfPredicate(field='Country', oneOf=['Afghanistan','India','Azerbaijan','
).add_params(selectYear).transform_filter(selectYear
).configure_title(
    anchor='middle',
    fontSize = 15
).configure_axis(
    labelFontSize=12,
     titleFontSize=12
).configure_legend(
    labelFontSize=12,
    titleFontSize =12,
    strokeColor='gray',
    fillColor='#EEEEEE',
    padding=5,
    cornerRadius=10,
    orient='bottom-right'
).configure_view(stroke=None)
employment_sector
```

Out[20]:

- Agriculture : Most South Asian women (about 60%) are employed in the field of agriculture and less

than 1% of women from North America region are employed in Agriculture.
- Industry : Between 8-20% of women from these regions are employed in the industry field.
- Service : A whopping 90% of women from North America are employed in the Service field.
- Overall, except for South Asian women, most women over the world are employed mostly in service fields.

Female labor force participation is one of the key drivers in the country's economic development. The visual on top shows the percentage of women's share in 2016 by each employment sector for five countries. The series of smaller bar plots show the same, between males and females, in each industry, country-wise. These sectors are gender-disaggregated data and are a broad classification from the world data bank.

The stacked bar plot visual is indicative that the agriculture sector in a developed nation like the United States shows minor percentages; less than 1% of females from the US are employed in Agriculture, whereas 90.86% of them are in the Service sectors.

This is indicative that the US imports more agriculture products while putting its workforce in service sectors. As with developing or least developed nations, more than 50% of women are in the Agriculture sector. Over the last 15 years, this trend has been different for each of these countries, mainly influenced by economic and political factors.

The industry sector includes occupations requiring more physical strength; evidently, percentages of males are more in this sector. Female percentage shares in the service sector have improved considerably for developing nations, while the US still tops over the years. As it is a well-developed nation, opportunities given to women in the employment sector seem fair.

What is the share of women in Parliament seats?

Exploring Parliament dataset

In [21]:
```
parliament.head()
```

Out[21]:

	Year	Azerbaijan	Afghanistan	India	Yemen, Rep.	United States	World
0	2020	17.355372	27.016129	14.364641	0.332226	27.464789	25.580431
1	2019	16.806723	27.868852	14.391144	0.332226	23.433875	24.636604
2	2018	16.800000	NaN	11.808118	0.000000	23.502304	24.097878
3	2017	16.800000	27.710843	11.808118	0.000000	19.354839	23.590337
4	2016	16.800000	27.710843	11.970534	0.000000	19.168591	23.091367

In [22]:
```
line = alt.Chart(parliament).mark_line(point=True).transform_fold(
     ['Azerbaijan','United States','India','Afghanistan','World']).encode(
    alt.X('Year:N', stack=None),
    alt.Y('value:Q',
          impute=alt.ImputeParams(method='mean'),
          axis=alt.Axis(tickMinStep = 5),
          scale=alt.Scale(domain=[0,30]),
          title = '% of Women in Parliament'),
    alt.Color('key:N'),
    alt.Tooltip('value:Q')
).properties(
    title ='Women % in Parliament over the years',
    width=700
)
```

Choosing heatmap

In [23]:
```
parl_hm = alt.Chart(parliament).mark_rect().transform_fold(
     ['Azerbaijan','United States','India','Afghanistan','World']).encode(
    alt.X('Year:N'),
    alt.Y('key:N',sort=['Afghanistan','India','Azerbaijan','United States','World'], tit
    alt.Color('value:Q',
              scale=alt.Scale(range=heatmap1),
              legend=alt.Legend(orient='right', titleOrient='top',
                                title='%')),
    tooltip= alt.Tooltip('value:Q', format='.1f')
    #alt.Size('value:Q')
).properties(
    width= 750,
    height=220,
    title ='Women Share(%) in Parliament over the years'
).transform_filter(
    'datum.Year > 2000'
).configure_title(
    anchor='middle',
    fontSize = 15
).configure_axis(
    labelFontSize=12,
    labelAngle=0,
     titleFontSize=12
).configure_legend(
    labelFontSize=9,
    titleFontSize =12,
    strokeColor='gray',
    fillColor='#EEEEEE',
    padding=5,
    cornerRadius=10,
    orient='bottom-right'
)
parl_hm
```

Out[23]:

As years progress, women are securing more seats in the parliament. However, the rise of the percentages in the last 20 years is only 11%, 14% (2001) to 25%(2020) world average.

Afghanistan has a higher proportion than the United States; this does not mean that Afghanistan is moving toward equal representation, but rather that the United States ranks below a nation with a high GI index.

Although more women in Afghanistan can hold seats in government parliament, this doesn't translate to power. Several other factors show that Afghan women are mistreated. Time will tell if the percentage reaches even 50% in these countries.

What percentage of parents return to the workforce after having a child?

In [24]:
```
labor_parent=lp[:4]
labor_parent = labor_parent.rename(columns={"Age of youngest child ":"child_age"})
```

In [25]:
```
labor_parent=pd.melt(labor_parent,id_vars=['child_age'],var_name='metrics', value_name='
labor_parent.head()
```

Out[25]:

	child_age	metrics	values

0	under 3 years	Mothers	63.3
1	3 to 5 years	Mothers	69.0
2	6 to 17 years	Mothers	75.4
3	under 18 years	Mothers	71.2
4	under 3 years	Fathers	93.5

In [26]:

```
parentperc = alt.Chart(labor_parent).mark_bar().encode(
    alt.Y('values:Q', title='Percent %'),
    x = alt.X("metrics:N", title=None, axis=None),
    color=alt.Color('metrics:N', scale=alt.Scale(range =heatmap), title='Parent'),
    tooltip = ['values'],
    column=alt.Column('child_age:N',title=("Percentage of Parent returning to Workforce
                      sort=["under 3 years", "3 to 5 years", "6 to 17 years","under 18 y
).transform_filter(
    'datum.child_age != "under 18 years"'
).properties(
    height = 400,
    width=150
).configure_axis(
    labelFontSize=12,
    titleFontSize=12
).configure_title(
    anchor='middle',
    fontSize = 15
).configure_header(
    titleFontSize=15,
    labelFontSize=12
).configure_legend(
    labelFontSize=10,
    titleFontSize =12,
    strokeColor='gray',
    fillColor='#EEEEEE',
    padding=5,
    cornerRadius=10,
    orient='right'
)

parentperc
```

Out[26]:

The goal of this visual is to address the issue of the unequal dedication of years for parenting. If both parents make the decision to have a child, does time for parenting lie evenly on the parent's shoulders?

It appears that women with younger-aged children are less likely to be in the work market, and as the child grows, they tend to return to the labor force. However, the presence of a new child does not affect men's careers, as the highest labor force is when the child is under age 3.

Is this fact being considered in the future when a woman has a career gap on her resume or is it being treated as a lack of career experience? In a world where balance is not maintained in child care, there should be balance in future opportunities.

In which occupations women are being paid more than men?

In [27]:

```
#Wage per Occupation Data Manipulation
occupation = pd.read_excel('wage_per_occupation.xlsx', sheet_name="Table 2")
occupation = occupation[3:]
```

```
data=occupation.reset_index()

data = data[4:]

data.columns = ['new_col1','Occupation', 'Number of workers/total', 'Median weekly earni
                'Standard error of median/total', 'Number of workers/women',
                'Median weekly earnings/women', 'Standard error of median/women',
               'Number of workers/men','Median weekly earnings/men','Standard error of m
                "Women's earnings as a percentage of men's"]
data = data.reset_index()
data = data.drop(columns=['new_col1'])

occup_data = pd.wide_to_long(data,
                             stubnames=['Number of workers', 'Median weekly earnings','S
                             i='index', j='group',
                             sep='/', suffix=r'\w+')
occup_data = occup_data.reset_index()

occup_data = occup_data.drop(columns=['index'])

occup_data = occup_data.rename(columns={"Women's earnings as a percentage of men's":'wom
                            "Occupation":"occupation",
                            "Number of workers":'num_work',
                            "Median weekly earnings":'median_week_earn',
                            "Standard error of median":'std_error_med'})

# filter missing/invalid values
occup_data = occup_data[(occup_data['women_earn_percentage'] != '-') & (occup_data['grou

occup_data.fillna(value = -1, inplace = True)

occup_data = occup_data[(occup_data['occupation']!= -1) & (occup_data['median_week_earn'
```

In [28]:

```
occup_data
```

Out[28]:

	group	women_earn_percentage	occupation	num_work	median_week_earn	std_error_med
598	women	73.8	Management, professional, and related occupations	25933	1164	4
599	women	76.4	Management, business, and financial operations...	9729	1274	12
600	women	77.5	Management occupations	5747	1347	12
601	women	75.6	Chief executives	363	2051	91
602	women	80.5	General and operations managers	281	1241	30
...	...	...	...	...	...	...
1763	men	102.2	Bus drivers, transit and intercity	89	774	54
1764	men	72.7	Driver/sales workers and truck drivers	2409	916	14
1783	men	88.5	Laborers and freight, stock, and material move...	1268	672	9
1785	men	90.1	Packers and packagers, hand	205	604	8

1786	men	95.7	Stockers and order fillers	714	602	8

298 rows × 6 columns

In [29]:

```
# Wage Gap Bar Chart
bar_chart = alt.Chart(occup_data).mark_bar().transform_calculate(
    wage_gap = 'datum.women_earn_percentage - 100',
    gender_high_pay = 'datum.wage_gap > 0 ? "women earn more": "men earn more"'
).encode(
    x=alt.X("occupation:N", title ='Occupation', axis = None),
    y=alt.Y("wage_gap:Q",title ='Wage gap in %'),
    tooltip = ['occupation','women_earn_percentage'],
    color=alt.Color('gender_high_pay:N', scale=alt.Scale(range =heatmap), title=None)

).properties(title = 'Women Wage Gap per Occupation',width=1000)




bar_chart_wage_gap = bar_chart.properties(
    height = 400,
    width=900
).configure_axis(
    labelFontSize=12,
    titleFontSize=12
).configure_title(
    anchor='middle',
    fontSize = 15
).configure_header(
    titleFontSize=15,
    labelFontSize=12
).configure_legend(
    labelFontSize=10,
    titleFontSize =12,
    strokeColor='gray',
    fillColor='#EEEEEE',
    padding=5,
    cornerRadius=10,
    orient='bottom-right'
)
bar_chart_wage_gap
```

Out[29]:

These visual carries a huge message, as women have higher paychecks only in 5 out of 149 occupations, and the following are the list of those occupations: Bus Drivers, Fast food and counter workers, Office and Administrative workers, producers and directors, and Wholesale and Retail buyers.

The highest wage gap is seen in the Legal occupations field, which is one of the highest-paid occupations. The height of bar charts where women are getting paid more is significantly less than of opposite ones. This means that even if women are paid more in those occupations, the difference in pay is not that huge. This visual carries fair analysis since the median earnings were classified by each occupation

What is the Adolescent Fertility Rate and Maternal Mortality rate?

Can there be any relation for factors with enrolment of women into secondary Education?

In [30]:

```
#color palette list
```

```
color_5_category =['#3A2A51','#FF7075' ,"#FFD35C",'#52A675',"#FFADB0"] #3 distinct
W = 430
sort_cty=['Yemen, Rep.','Afghanistan','India','Azerbaijan','United States']
```

In [31]:

```
# filter by country

jobs  = pd.read_csv("JobsData.csv")
inequality  = pd.read_csv("gender-inequality-index-from-the-human-development-report.csv

inequality_cty =inequality[inequality["Entity"].isin(["India","United States"
                                          ,"Yemen, Rep."
                                                   ,"Afghanistan"
                                              ,"Azerbaijan"
                                             ])]
```

In [32]:

```
inequality_2005_2021 = inequality_cty[inequality_cty["Year"]>= 2005]

inequality_2021 =  inequality_cty[inequality_cty["Year"]== 2021]
inequality_2021
inequality_world_2021 = inequality[inequality["Year"]== 2021]
```

In [33]:

```
world = geopandas.read_file(geopandas.datasets.get_path('naturalearth_lowres'))
world.head()
```

```
C:\Users\rrads\AppData\Local\Temp\ipykernel_24608\850357869.py:1: FutureWarning: The geo
pandas.dataset module is deprecated and will be removed in GeoPandas 1.0. You can get th
e original 'naturalearth_lowres' data from https://www.naturalearthdata.com/downloads/11
0m-cultural-vectors/.
  world = geopandas.read_file(geopandas.datasets.get_path('naturalearth_lowres'))
```

Out[33]:

	pop_est	continent	name	iso_a3	gdp_md_est	geometry
0	889953.0	Oceania	Fiji	FJI	5496	MULTIPOLYGON (((180.00000 -16.06713, 180.00000...
1	58005463.0	Africa	Tanzania	TZA	63177	POLYGON ((33.90371 -0.95000, 34.07262 -1.05982...
2	603253.0	Africa	W. Sahara	ESH	907	POLYGON ((-8.66559 27.65643, -8.66512 27.58948...
3	37589262.0	North America	Canada	CAN	1736425	MULTIPOLYGON (((-122.84000 49.00000, -122.9742...
4	328239523.0	North America	United States of America	USA	21433226	MULTIPOLYGON (((-122.84000 49.00000, -120.0000...

In [34]:

```
merge_DF = pd.merge(world, inequality_world_2021, left_on='iso_a3', right_on='Code')
merge_DF.columns =['pop_est', 'continent', 'name', 'iso_a3', 'gdp_md_est', 'geometry',
       'Entity', 'Code', 'Year',
       'GDI']

merge_DF
```

Out[34]:

	pop_est	continent	name	iso_a3	gdp_md_est	geometry	Entity	Code	Year	GDI
0	889953.0	Oceania	Fiji	FJI	5496	MULTIPOLYGON (((180.00000 -16.06713, 180.00000...	Fiji	FJI	2021	0.318
1	58005463.0	Africa	Tanzania	TZA	63177	POLYGON ((33.90371 -0.95000,	Tanzania	TZA	2021	0.560

						34.07262 -1.05982...				
2	37589262.0	North America	Canada	CAN	1736425	MULTIPOLYGON (((-122.84000 49.00000, -122.9742...	Canada	CAN	2021	0.069
3	328239523.0	North America	United States of America	USA	21433226	MULTIPOLYGON (((-122.84000 49.00000, -120.0000...	United States	USA	2021	0.179
4	18513930.0	Asia	Kazakhstan	KAZ	181665	POLYGON ((87.35997 49.21498, 86.59878 48.54918...	Kazakhstan	KAZ	2021	0.161
...	...	...	...	...	...	...	...	...	...	...
153	2083459.0	Europe	North Macedonia	MKD	12547	POLYGON ((22.38053 42.32026, 22.88137 41.99930...	North Macedonia	MKD	2021	0.134
154	6944975.0	Europe	Serbia	SRB	51475	POLYGON ((18.82982 45.90887, 18.82984 45.90888...	Serbia	SRB	2021	0.131
155	622137.0	Europe	Montenegro	MNE	5542	POLYGON ((20.07070 42.58863, 19.80161 42.50009...	Montenegro	MNE	2021	0.119
156	1394973.0	North America	Trinidad and Tobago	TTO	24269	POLYGON ((-61.68000 10.76000, -61.10500 10.890...	Trinidad and Tobago	TTO	2021	0.344
157	11062113.0	Africa	S. Sudan	SSD	11998	POLYGON ((30.83385 3.50917, 29.95350 4.17370, ...	South Sudan	SSD	2021	0.587

158 rows × 10 columns

In [35]:

```
GDI_Trend =  ( alt.Chart(inequality_2005_2021).mark_line(
).encode(
    alt.X("Year:N"  )
    ,alt.Y( "Gender Inequality Index:Q")
#     ,column = "Name:N"
#     longitude='longitude:Q', # apply the field named 'longitude' to the longitude chan
#     latitude='latitude:Q'    # apply the field named 'latitude' to the latitude channe
    ,color  =  alt.Color("Entity:N"
    ,  scale = alt.Scale(range = color_5_category)
                       ,sort =  sort_cty)
#     , tooltip = ["name" , "GDI"]
)).properties(
    width=W,
#     / height=500
    title  = "Gender Inequality Index"
)
```

In [36]:

```
GDI_bar =  ( alt.Chart(inequality_2021).mark_bar(
).encode(
    alt.X("Entity:N" ,sort =  sort_cty )
    ,alt.Y( "Gender Inequality Index:Q")
#     ,column = "Name:N"
#     longitude='longitude:Q', # apply the field named 'longitude' to the longitude chan
#     latitude='latitude:Q'    # apply the field named 'latitude' to the latitude channe
    ,color  =  alt.Color("Entity:N"
    ,  scale = alt.Scale(range = color_5_category)
                       ,sort =  sort_cty
                        ,  legend=alt.Legend(orient='top', titleOrient='left',
                             title='Country'
                  ))
#     , tooltip = ["name" , "GDI"]
)).properties(
    width=W,
#    / height=500
    title  = "Gender Inequality Index - 2021"
)
```

In [37]:

```
(GDI_Trend | GDI_bar).configure_view(
    stroke=None
).configure_legend(
    labelFontSize=12,
    titleFontSize =12,
    strokeColor='gray',
    fillColor='#EEEEEE',
    padding=5,
    cornerRadius=10,
    orient='top-right'
)
```

Out[37]:

In [38]:

```
Jobs = jobs[['Country Name', 'Country Code', 'Indicator Name', 'Indicator Code',
        '2011']]
Jobs.columns = Jobs.columns.astype(str)
Stats5countries =  Jobs[Jobs["Country Name"].isin(sort_cty)]
```

In [39]:

```
Female_secodary_enrolment = Stats5countries[Stats5countries["Indicator Code"].isin([
                                                                         "SE.SEC.ENRR.FE
```

In [40]:

```
Secondary_bar = alt.Chart(Female_secodary_enrolment).mark_line( stroke = "#65605D"  , co
alt.X("Country Name:N", title = None ,sort=sort_cty, axis=alt.Axis(labels=False))
, alt.Y("2011:Q" , title = "School enrollment, secondary, female (% gross)" , scale=alt.
)
```

In [41]:

```
Fertility  = pd.read_csv("Adolescent_fertilirt.csv")
Fertility_2017 = Fertility[Fertility["Year"] == 2017]
Fertility_2017
```

Out[41]:

	Year	Adolescent fertility rate (births per 1,000 women ages 15-19)	Country
3	2017	55.838	Azerbaijan
24	2017	68.957	Afghanistan
45	2017	60.352	Yemen, Rep.
66	2017	13.177	India

87	2017	19.860	United States

In [42]:

```
fertility_bar = alt.Chart(Fertility_2017).mark_bar().encode(
alt.X("Country:N", title = None ,sort=sort_cty)
, alt.Y("Adolescent fertility rate (births per 1,000 women ages 15-19):Q"
        , title = "Adolescent fertility rate" )
    ,alt.Color("Country:N"  )
 ).transform_filter("datum.Country != 'World'").properties(width =W , title = "Adolescen


P1 = (fertility_bar + Secondary_bar.encode(
alt.Y("2011:Q" ,title = None , axis=alt.Axis(labels=False)))).resolve_scale(
    y="independent"
    , x = "independent"
).properties(width =W
            )
```

In [43]:

```
mortality  = pd.read_csv("Maternal_Mortality_ratio.csv")
mortality_2017 = mortality[mortality["Year"] == 2017]
mortality_2017
```

Out[43]:

	Year	Country	Maternal mortality ratio (per 100,000 live births)
0	2017	World	211
18	2017	Afghanistan	638
36	2017	Azerbaijan	26
54	2017	India	145
72	2017	Yemen, Rep.	164
90	2017	United States	19

In [44]:

```
mortality_bar = alt.Chart(mortality_2017).mark_bar().encode(
alt.X("Country:N", title = None ,sort=sort_cty )
, alt.Y("Maternal mortality ratio (per 100,000 live births):Q"  , title = "Maternal mort
    ,alt.Color("Country:N" , legend = None , scale = alt.Scale(range = color_5_category)
).transform_filter("datum.Country != 'World'").properties(width =W , title = "Maternal m
mortality_bar

p2 = (mortality_bar + Secondary_bar
     ).resolve_scale(
    y="independent"
    , x = "independent"
).properties(width =W , title = "Maternal mortality ratio (per 100,000 live births) - 20
```

In [45]:

```
mortality_trend = alt.Chart(mortality).mark_line().encode(
alt.X("Year:N")
,alt.Y("Maternal mortality ratio (per 100,000 live births)" , title ="Maternal mortality
,alt.Color("Country"
                   ,  scale = alt.Scale(range = color_5_category)
           , legend=alt.Legend(orient='top', titleOrient='left',
                               title='Country'
                  ))).transform_filter("datum.Country != 'World'"
                                  ).properties(width =W , title = "Trend of Maternal m
```

In [46]:

```
Fertility_trend = alt.Chart(Fertility).mark_line(
).encode(
alt.X("Year:N")
,alt.Y("Adolescent fertility rate (births per 1,000 women ages 15-19)"
       , title = "Adolescent fertility rate")
```

```
    ,alt.Color("Country")
).transform_filter("datum.Country != 'World'"
                  ).transform_filter("datum.Year <='2017'"
                                    ).properties(width =W
                                                 , title ="Trend of Adolescent fertility
```

In [47]:

```
Ferti_Mortalilty = (( Fertility_trend | mortality_trend) & (P1 | p2)
).resolve_scale(color = "independent").configure_legend(
    labelFontSize=12,
    titleFontSize =12,
    strokeColor='gray',
    fillColor='#EEEEEE',
    padding=5,
    cornerRadius=10,
    orient='top-right'
).configure_axis(
    labelFontSize=10,
     titleFontSize=10
    ,labelAngle=0
).configure_title(
    anchor='middle',
    fontSize = 12
)

Ferti_Mortalilty
```

Out[47]:

Summary of all visuals

What is the share of women employment by sectors?

In [48]:

```
employment_sector
```

Out[48]:

What is the share of women in Parliament seats?

In [49]:

```
parl_hm
```

Out[49]:

What is the Adolescent Fertility Rate and Maternal Mortality rate ? Can there be any relation for factors with enrolment of women into secondary Education?

In [50]:

```
Ferti_Mortalilty
```

Out[50]:

What percentage of parents return to the workforce after having a child?

In [51]:

```
parentperc
```

Out[51]:

In which occupations women are being paid more than men?

```
In [52]: bar_chart_wage_gap
```

```
Out[52]:
```

Conclusion

The project aimed to explore the main aspects driving the Gender inequality index. Some explored questions included factors such as mortality ratio, school enrollment of females, fertility rate, women in parliament, women returning to work after a child, women in employment sectors, and wage differences between genders. Some findings from our exploration;

Secondary education provided for females can lead to improvement in terms of maternal mortality and adolescent fertility. Wage difference analysis suggests that women sacrifice their careers and dedicate time to childcare, whereas men’s employment trend stays almost unaffected.

Furthermore, analysis of the trend of high-earning males in certain occupations remains unchanged, and only 3-4% of occupations for women are paid higher than men. Strengthening the collective power of women in leadership is perhaps the answer to bridging gaps.

The world has an average of 25% women’s share in parliament seats. It is a hopeful sign that there will be an increase in the coming years, and the world will move towards lower disparities between genders.

```
In [ ]:
```