Data Visualization

For Data Scientists using Python

USING OPEN SOURCED TOOLS

Nikhil Widhani
Data Scientist Environment and Climate Change Canada

Agenda

01 **Problem Statement**

Understand the need for data visualization for data scientists.

02 **Prerequisites**

Grab knowledge of essential Python libraries which help us to deal with data know which libraries are available in the market

03 **Tutorial**

Understanding Dash callbacks and Plotly charts,

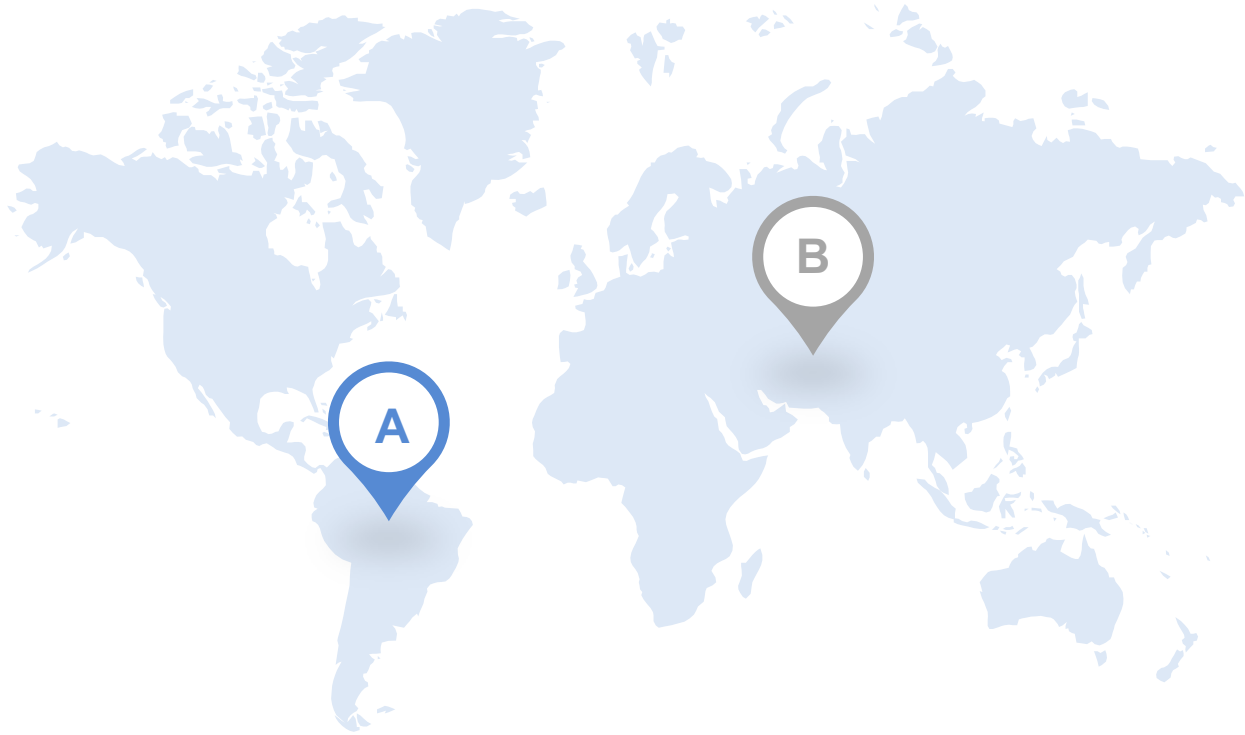
04 **Demo**

Demo based on real project done at ECCC.

Problem Statement

Why we need
Visualization ?

- Data visualization provides us with a quick, clear understanding of the information.
- Helps to identify trends and gain insights quickly.
- Helps to analyze at various levels of details
- Helps to tell our data story



PREREQUISITE
LIBRARIES

Data Structuring

- Pandas : For data manipulation and analysis
- NumPy : For Mathematical Functions

Data Visualization

- Matplotlib : For basic plotting understanding
- Plotly : Advanced Python Plotting
- Dash : For integrating Plotly plots to frontend applications

Tutorial - Dash Callbacks

```python
import dash
import dash_core_components as dcc
import dash_html_components as html
from dash.dependencies import Input, Output

external_stylesheets = ['https://codepen.io/chriddyp/pen/bWLwgP.css']

app = dash.Dash(__name__, external_stylesheets=external_stylesheets)

app.layout = html.Div([
    dcc.Input(id='my-id', value='initial value', type='text'),
    html.Div(id='my-div')
])


@app.callback(
    Output('my-div', 'children'),
    [Input('my-id', 'value')]
)
def update_output_div(input_value):
    return 'You\'ve entered "{}"'.format(input_value)


if __name__ == '__main__':
    app.run_server(debug=True)
```

1. Import necessary libraries

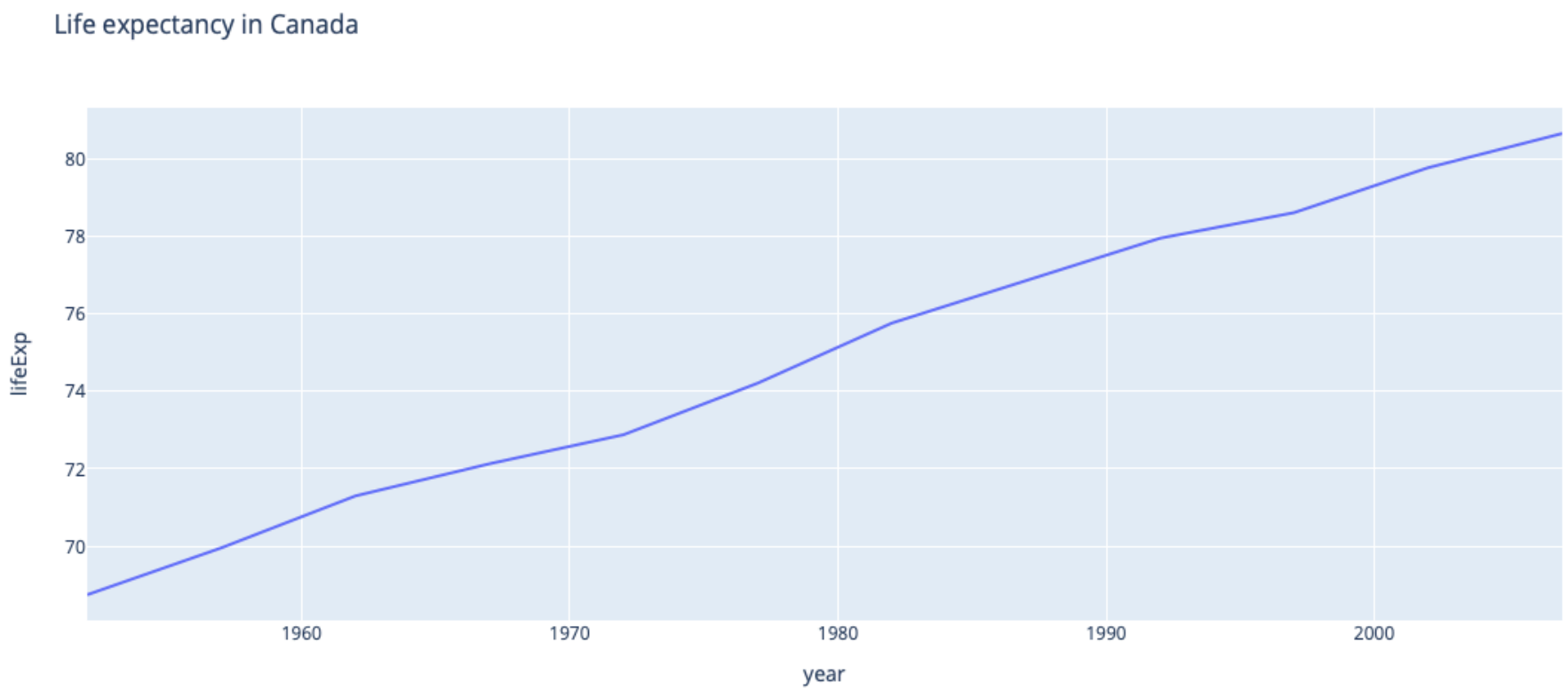
2. Define the app layout

3. Program the callbacks

Tutorial – Plotly Charts

```
import plotly.express as px

df = px.data.gapminder().query("country=='Canada'")
fig = px.line(df, x="year", y="lifeExp", title='Life expectancy in Canada')
fig.show()
```



Tutorial – Plotly with Dash

```
30  import dash
31  import dash_core_components as dcc
32  import dash_html_components as html
33  from dash.dependencies import Input, Output
34  import plotly.express as px
35
36  external_stylesheets = ['https://codepen.io/chriddyp/pen/bWLwgP.css']
37
38  app = dash.Dash(__name__, external_stylesheets=external_stylesheets)
39
40  app.layout = html.Div([
41      dcc.Store(id="my-id"),
42      dcc.Graph(id='my-div')
43  ])
44
45
46  @app.callback(
47      Output('my-div', 'figure'),
48      [Input('my-id', 'value')]
49  )
50  def update_output_div(input_value):
51      df = px.data.gapminder().query("country=='Canada'")
52      fig = px.line(df, x="year", y="lifeExp", title='Life expectancy in Canada')
53      return fig
54
55
56  if __name__ == '__main__':
57      app.run_server(debug=True)
```

1. Add dcc.Graph component

2. Return Figure to show the plot

PROJECT DEMO

Problem: To create a ML algorithm to predict company noncompliance.

Result: You used complex data science techniques to predict the noncompliance percentage based on historical data and generate results in a csv format.

name	address	city	province	postalcode	latitude	longitude	days_since_last_visit	ensemble_pred_prob	times_noncomplient
Foundation for International Training	7181 WOODBINE AVEN	MARKHAM	ON	L3R1A3	43.819131	-79.348759	311	0.432577101	4
COLEPRO PRODUCTIONS INC.	16 DEVON AVENUE	WESTMOUNT	QC	H3Y1A2	45.490515	-73.615388	224	0.100768345	3
HOPITAL VETERINAIRE DE COTE ST. LU	7930 BAILY ROAD	COTE ST. LUC	QC	H4W3J8	45.460747	-73.663142	262	0.213657525	2
BEN W. CROW & ASSOCIES DU CANADA	250 CONSUMERS RO/	WILLOWDALE	ON	M2J4V6	43.772028	-79.330594	89	0.507619841	2
SERVICES DE DISQUES DELTA INC.	2379 GUENETTE ST.	ST. LAURENT	QC	H4R2E9	45.499892	-73.728608	232	0.310690445	3
PROTECTION INCENDIE DESCO LTEE	RR 2, 4E LIGNE	NAVAN	ON	K4B1H9	45.428012	-75.470078	328	0.373741775	2
80922 CANADA LTD.	80 MUNRO BLVD.	TORONTO	ON	M2P1C4	43.746768	-79.394519	106	0.335185149	2
EMI TECHNOLOGY LTD.	FIRST CANADIAN PLAC	TORONTO	ON	M5X1B2	43.665614	-79.347717	325	0.742791886	3

Demo

Useful Links

Dash Documentation: https://dash.plotly.com

Dash App Gallery: https://dash-gallery.plotly.host/Portal/

Plotly examples: https://plotly.com/python/

Detailed Dash Tutorial: https://towardsdatascience.com/how-to-build-a-complex-reporting-dashboard-using-dash-and-plotl-4f4257c18a7f

Advanced Techniques and Libraries:

- Django
- Java Script Chart libraries (Apex Charts, Chart JS)

Thank You

Future Data Scientists ;)