Assignment – Day 17

-Sarthak Niranjan Kulkarni (Maverick)

- sarthakkul2311@gmail.com - (+91) 93256 02791

28/11/2024 (Thursday)

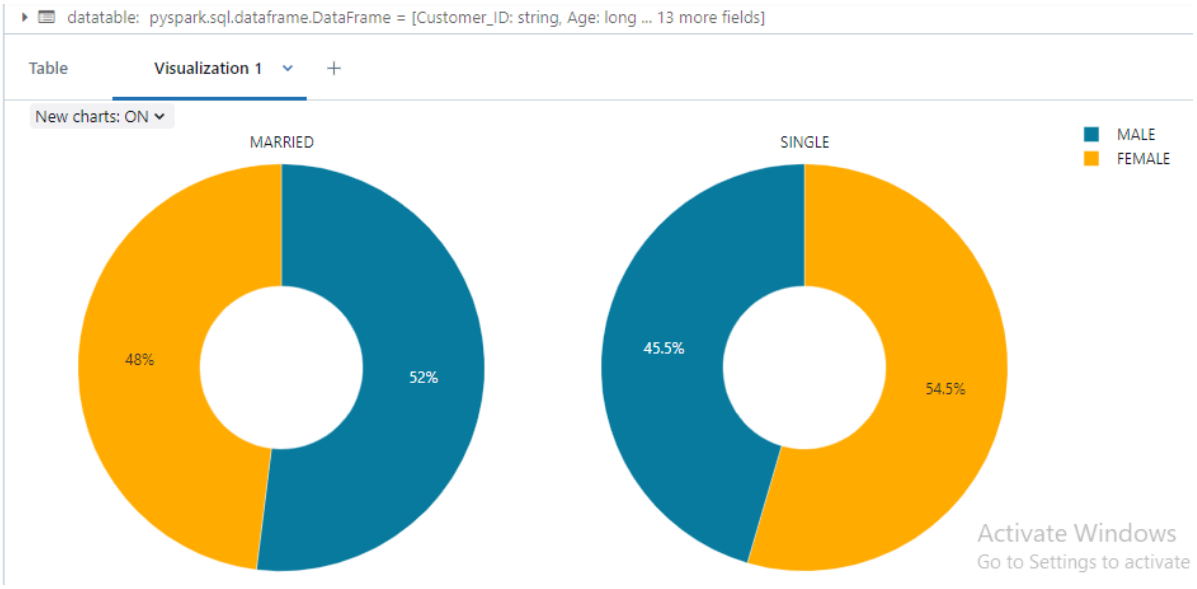
Practice on Visualization: -

1. Loading and Displaying Data from the 'loan' Table

→

Table | Visualization 1 +

	Customer_ID	Age	Gender	Occupation	Marital Status	Family Size	Income	Expenditure
1	IB14001	30	MALE	BANK MANAGER	SINGLE	4	50000	2
2	IB14008	44	MALE	PROFESSOR	MARRIED	6	51000	1
3	IB14012	30	FEMALE	DENTIST	SINGLE	3	58450	2
4	IB14018	29	MALE	TEACHER	MARRIED	5	45767	1
5	IB14022	34	MALE	POLICE	SINGLE	4	43521	1
6	IB14024	55	FEMALE	NURSE	MARRIED	6	34999	1
7	IB14025	39	FEMALE	TEACHER	MARRIED	6	46619	1
8	IB14027	51	MALE	SYSTEM MANAGER	MARRIED	3	49999	1
9	IB14029	24	FEMALE	TEACHER	SINGLE	3	45008	1
10	IB14031	37	FEMALE	SOFTWARE ENGINEER	MARRIED	5	55999	2
11	IB14032	24	MALE	DATA ANALYST	SINGLE	4	60111	2
12	IB14034	32	MALE	PRODUCT ENGINEER	MARRIED	6	null	2
13	IB14037	54	FEMALE	TEACHER	MARRIED	5	48099	1
14	IB14039	45	MALE	ACCOUNT MANAGER	MARRIED	7	45777	1



2. Loading and Displaying Data from 'export' Table and Delta Location

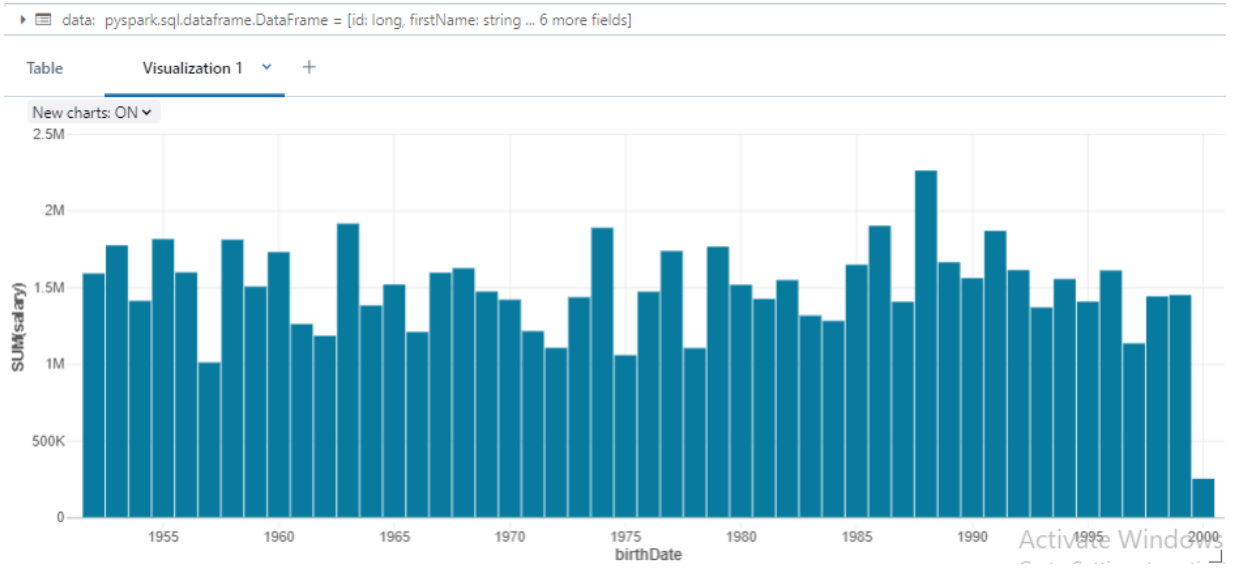
→ spark.table("default.export")

data = spark.read.format("delta").load("dbfs:/user/hive/warehouse/export")

data.display()

Table | Visualization 1 +

	id	firstName	middleName	lastName	gender	birthDate	ssn	salary
1	1	Pennie	Carry	Hirschmann	F	1955-07-02T04:00:00.000+00:...	981-43-9345	56172
2	2	An	Amira	Cowper	F	1992-02-08T05:00:00.000+00:...	978-97-8086	40203
3	3	Quyen	Marlen	Dome	F	1970-10-11T04:00:00.000+00:...	957-57-8246	53417
4	4	Coralie	Antonina	Marshal	F	1990-04-11T04:00:00.000+00:...	963-39-4885	94727
5	5	Terrie	Wava	Bonar	F	1980-01-16T05:00:00.000+00:...	964-49-8051	79908
6	6	Chassidy	Concepcion	Bourthouloume	F	1990-11-24T05:00:00.000+00:...	954-59-9172	64652
7	7	Geri	Tambra	Mosby	F	1970-12-19T05:00:00.000+00:...	968-16-4020	38195
8	8	Patria	Nancy	Arstall	F	1985-01-02T05:00:00.000+00:...	984-76-3770	102053
9	9	Terese	Alfredia	Tocque	F	1967-11-17T05:00:00.000+00:...	967-48-7309	91294
10	10	Wava	Lyndsey	Jeandon	F	1963-12-30T05:00:00.000+00:...	997-82-2946	56521
11	11	Sophie	Emerita	Hearn	F	1979-09-17T04:00:00.000+00:...	977-66-4483	90920
12	12	Jodie	Tabetha	Laneham	F	1959-01-31T05:00:00.000+00:...	923-24-9769	90634
13	13	Marietta	Mandi	Yansons	F	1974-02-19T04:00:00.000+00:...	900-34-8083	93162
14	14	Caridad	Maire	Snelle	F	1960-09-26T04:00:00.000+00:...	992-11-7062	38859
15	15	Yasmine	Meg	Edworthye	F	1960-01-29T05:00:00.000+00:...	922-12-9862	76220



Summary on Visualization: -

In Azure Databricks, data visualization using PySpark can help you easily interpret and present data insights. PySpark allows you to work with large datasets and perform complex transformations before visualizing the results. The display() function in Databricks provides a powerful way to visualize DataFrames directly in the notebook interface. When you load data into PySpark, whether from a table or a Delta file, you can quickly visualize it using Databricks' built-in visualization tools.

Visualizations like bar charts, line graphs, and scatter plots can be created with just a few clicks, providing an intuitive way to explore data patterns. You can create custom visualizations to examine trends over time, compare categories, or understand distributions. Databricks also supports interactive visualization, which means you can drill down into the data, filter values, and adjust axes for better clarity.

By using PySpark for data processing and Databricks for visualization, you can enhance your data exploration experience without switching tools or environments. This integration makes it easier to share insights with others and perform interactive analysis in real-time. Visualizations are crucial for communicating data findings in an understandable and impactful way.