

## QUERY PROCESSING - LAB PROGRAMS

1. Write a Pandas program to select distinct department id from employees file.

DEPARTMENT_ID	DEPARTMENT_NAME	MANAGER_ID	LOCATION_ID
10	Administration	200	1700
20	Marketing	201	1800
30	Purchasing	114	1700
40	Human Resources	203	2400
50	Shipping	121	1500
60	IT	103	1400
70	Public Relations	204	2700
80	Sales	145	2500
90	Executive	100	1700
100	Finance	108	1700
110	Accounting	205	1700
120	Treasury	0	1700
130	Corporate Tax	0	1700
140	Control And Credit	0	1700
150	Shareholder Services	0	1700
160	Benefits	0	1700
170	Manufacturing	0	1700
180	Construction	0	1700
190	Contracting	0	1700
200	Operations	0	1700
210	IT Support	0	1700
220	NOC	0	1700
230	IT Helpdesk	0	1700
240	Government Sales	0	1700
250	Retail Sales	0	1700
260	Recruiting	0	1700
270	Payroll	0	1700

**CODE & OUTPUT:**

The screenshot shows a Jupyter Notebook with two cells. The first cell contains Python code to create a DataFrame with 19 rows and 3 columns: DEPARTMENT\_ID, MANAGER\_ID, and LOCATION\_ID. The second cell displays the output of the DataFrame, showing the first 19 rows of data.

```

import pandas as pd
# Create the DataFrame
data = {
    "DEPARTMENT_ID": [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260],
    "DEPARTMENT_NAME": [
        "Administration", "Marketing", "Purchasing", "Human Resources", "Shipping",
        "Public Relations", "Sales", "Executive", "Finance", "Accounting", "Treasury",
        "Corporate Tax", "Control And Credit", "Shareholder Services", "Benefits",
        "Manufacturing", "Construction", "Contracting", "Operations", "IT Support",
        "NOC", "IT Helpdesk", "Government Sales", "Retail Sales", "Recruiting", "Facilities"
    ],
    "MANAGER_ID": [200, 201, 114, 203, 121, 103, 204, 100, 108, 205, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
    "LOCATION_ID": [1700, 1800, 1800, 1700, 2400, 1500, 1400, 2700, 2500, 1700, 1700, 1700, 1700, 1700, 1700, 1700, 1700, 1700, 1700, 1700, 1700, 1700, 1700, 1700, 1700, 1700]
}
df = pd.DataFrame(data)
dept_id= df["DEPARTMENT_ID"].drop_duplicates()
print(dept_id)

```

The output of the DataFrame is as follows:

	DEPARTMENT_ID	DEPARTMENT_NAME	MANAGER_ID	LOCATION_ID
0	10	Administration	200	1700
1	20	Marketing	201	1800
2	30	Purchasing	114	1800
3	40	Human Resources	203	1700
4	50	Shipping	121	1700
5	60	Public Relations	103	1700
6	70	Sales	204	1700
7	80	Executive	100	1700
8	90	Finance	108	1700
9	100	Accounting	205	1700
10	110	Treasury	0	1700
11	120	Corporate Tax	0	1700
12	130	Control And Credit	0	1700
13	140	Shareholder Services	0	1700
14	150	Benefits	0	1700
15	160	Manufacturing	0	1700
16	170	Construction	0	1700
17	180	Contracting	0	1700
18	190	Operations	0	1700