INTERNSHIP PROJECT PHASE-DEPLOYMENT

PROJECT 6 – DEVELOPMENT AND TESTING OF YELP DATA ANALYSIS PROJECT ON IBM CLOUD

Deployment Steps

Step 1: IBM Cloud Setup

Login to IBM Cloud.

Navigate to the Catalog and provision an Object Storage instance.

Create a bucket named yelp-data-bucket and upload both JSON datasets.

Step 2: Watson Studio Project Setup

Go to Watson Studio and create a new project.

Under Storage, link your previously created Object Storage bucket.

Add collaborators if working in a team.

Step 3: Notebook Deployment

From the Watson Studio dashboard, click Add to Project > Notebook.

Import the existing Data_Analysis_on_Yelp_Dataset.ipynb notebook.

Ensure the runtime environment has Python 3.9 and the necessary packages (pandas, seaborn, matplotlib).

Step 4: Data Access Configuration

Use project_lib or ibm_boto3 to access Object Storage within the notebook.

```
from project_lib import Project
project = Project.access()
file_path = project.get_file("yelp_academic_dataset_business.json")
```

Step 5: Execution

Execute notebook cells sequentially.

Save outputs as images or CSVs to Object Storage.

Use project.save_data() for persistence.

Step 6: Dashboard Integration (Optional)

Use IBM Cognos Dashboard Embedded to create visual dashboards.

Import CSV results or connect directly to Object Storage.

Post-Deployment Activities

- Testing: Validate notebook execution and output generation.
- Logging: Use print() statements and error handling for debugging.
- Security: Manage IAM roles and access controls for Object Storage and Watson Studio.
- Backup: Schedule regular backup of final outputs.

Automation (Optional)

- You can deploy automation using IBM Cloud Functions and IBM Schematics:
- Schedule notebooks using Cloud Functions and REST APIs.
- Trigger analysis on dataset upload using Cloud Object Storage events.
- Use webhooks for notification on task completion.

Sample CLI Command:

ibmcloud fn action create runNotebook --kind python:3 runNotebook.py

Deployment challenges and solutions:

- **Access token issues**: Use project tokens or IAM roles properly to ensure secure and authorized access.
- **Runtime limitations**: Use the Lite or Standard Watson Studio plan with appropriate resource allocation.
- **Slow loading of large JSON**: Optimize loading by using pandas.read_json() with lines=True.

Maintenance and Monitoring

- Regularly check Object Storage for space and data integrity.
- Update dependencies and runtime environments as needed.
- Periodically evaluate performance metrics.
- Use pip freeze > requirements.txt to maintain environment reproducibility.