Kedro Project Setup Guide: confidence_score

Complete Steps to Create Kedro Project with 4 Empty Nodes

1. Prep Environment (Windows PowerShell)
```
mkdir confidence_score
cd confidence_score
python -m venv .venv
.venv\Scripts\Activate.ps1
pip install "kedro>=0.18" "kedro-datasets[pandas]" kedro-viz pandas
```

2. Create Kedro Project
```
kedro new
```
- Project name: confidence_score
- Repository name: confidence_score
- Python package name: confidence_score

```
cd confidence_score
```

3. Create Pipeline
```
kedro pipeline create confidence
```

4. catalog.yml
conf/base/catalog.yml:
```
raw_data:
  type: kedro_datasets.pandas.CSVDataset
  filepath: data/01_raw/raw_data.csv
  save_args:
    index: False

predictions:
  type: kedro_datasets.pandas.CSVDataset
  filepath: data/02_intermediate/predictions.csv
```

```yaml
  save_args:
    index: False


metrics:
  type: kedro_datasets.json.JSONDataset
  filepath: data/03_primary/metrics.json
```


5. nodes.py
src/confidence_score/pipelines/confidence/nodes.py:
```python
import pandas as pd
from typing import Dict, Any

def load_data() -> pd.DataFrame:
    print(">> RUN: load_data")
    return pd.DataFrame(columns=["id", "value"])

def run_inference(raw_data: pd.DataFrame) -> pd.DataFrame:
    print(">> RUN: run_inference")
    if raw_data is None or raw_data.empty:
        return pd.DataFrame(columns=["id", "prediction"])
    df = raw_data.copy()
    df["prediction"] = df.get("value", None)
    return df

def calculate_score(predictions: pd.DataFrame) -> Dict[str, Any]:
    print(">> RUN: calculate_score")
    if predictions is None or predictions.empty or "prediction" not in predictions.columns:
        return {"average_prediction": None}
    return {"average_prediction": float(predictions["prediction"].mean())}

def create_visualization(metrics: Dict[str, Any]) -> None:
    print(">> RUN: create_visualization ->", metrics)
    return None
```


6. pipeline.py
src/confidence_score/pipelines/confidence/pipeline.py:
```python
from kedro.pipeline import Pipeline, node
from . import nodes

def create_pipeline(**kwargs) -> Pipeline:
```

```
    return Pipeline([
        node(nodes.load_data, None, "raw_data", name="load_data"),
        node(nodes.run_inference, "raw_data", "predictions", name="run_inference"),
        node(nodes.calculate_score, "predictions", "metrics", name="calculate_score"),
        node(nodes.create_visualization, "metrics", None, name="create_visualization"),
    ])
```

7. pipeline_registry.py
src/confidence_score/pipeline_registry.py:
```
from typing import Dict
from kedro.pipeline import Pipeline
from confidence_score.pipelines.confidence.pipeline import create_pipeline

def register_pipelines() -> Dict[str, Pipeline]:
    confidence = create_pipeline()
    return {
        "__default__": confidence,
        "confidence": confidence,
    }
```

8. Create Folders
```
New-Item -ItemType Directory -Force -Path data\01_raw
New-Item -ItemType Directory -Force -Path data\02_intermediate
New-Item -ItemType Directory -Force -Path data\03_primary
```

9. Run Pipeline
```
kedro run --pipeline confidence --verbose
```

Expected prints:
```
>> RUN: load_data
>> RUN: run_inference
>> RUN: calculate_score
>> RUN: create_visualization -> {'average_prediction': None}
```

10. Visualize Pipeline

```
kedro viz
```

Open http://127.0.0.1:4141 in your browser.