Jedha

SuperCourier Projects

Practical Assessment for DE & DS Certifications

Jedha

The Business Context

- **SuperCourier**: Leading delivery company facing challenges with on-time performance
- **Business Goal**: Improve delivery prediction and reduce delays
- **Technical Approach**:
  - Build a data pipeline (DE project)
  - Create a prediction model (DS project)
- **The Challenge**: Complete these tasks under tight time constraints (2 hours each)

Jedha

Two Interconnected Projects

Data Engineering Project

- Create an ETL pipeline for delivery data
- Generate synthetic dataset from multiple sources
- Deliver clean, structured data ready for analysis
- **Time Frame**: 2 hours

Data Science Project

- Analyze the dataset produced by DEs
- Identify key factors affecting delivery delays
- Build a predictive model for delivery timing
- **Time Frame**: 2 hours

Jedha

Project 1: Data Engineering Challenge

The Challenge: Delivery Delay Pipeline

- Design and implement a data pipeline
- Extract from multiple sources
- Transform and enrich delivery data
- Load into a structured format ready for analysis
- Handle missing values and data quality issues

Jedha

DE Project - Certification Alignment

Key Tasks

- Create SQLite database with delivery records
- Generate weather and tracking data
- Join and transform data sources
- Calculate delivery metrics
- Export clean dataset with quality checks

Certification Blocks

RNCP37172BC01 Concevoir un projet d'architecture de gestion de données massives

RNCP37172BC02 Élaborer une solution technique de collecte et de traitement de données massives

RNCP37172BC03 Déployer l'architecture de gestion de données massives

Jedha

DE Project - Skills Assessment

Technical Skills

- Database design
- Data extraction techniques
- Transformation logic
- Data quality management
- Pipeline orchestration

Key Competencies

- **BC01**: Architecture design choices
- **BC02**: Data collection methodology
- **BC02**: Data transformation strategy
- **BC03**: Implementation and deployment
- **BC03**: Quality control mechanisms

Jedha

Project 2: Data Science Challenge

The Challenge: Delivery Delay Investigation

- Analyze delivery dataset
- Identify factors influencing delays
- Create visualizations of key patterns
- Build a predictive model
- Provide business recommendations

Jedha

DS Project - Certification Alignment

Key Tasks

- Perform exploratory data analysis
- Preprocess data for modeling
- Identify delay correlation factors
- Build classification model
- Evaluate model performance
- Communicate insights

Certification Blocks

RNCP38777BC01 Concevoir et piloter la gouvernance des données RNCP38777BC03 Concevoir et mettre en oeuvre des pipelines de données (pour l'IA) RNCP38777BC04 Construire déployer et piloter des solutions d'IA

Jedha

DS Project - Skills Assessment

Technical Skills

- Data exploration
- Statistical analysis
- Feature engineering
- Machine learning modeling
- Model evaluation
- Data visualization

Key Competencies

- **BC01**: Data analysis methodology
- **BC03**: Data preparation for ML
- **BC03**: Feature importance analysis
- **BC04**: Model selection and tuning
- **BC04**: Solution implementation
- **BC04**: Results interpretation

Jedha

The Data Connection

What DEs Provide

- Clean, structured dataset
- Calculated delivery times
- Enriched features (weather, etc.)
- Quality-checked data
- Documentation of data sources

What DSs Receive

- Ready-to-analyze dataset
- 2,000 delivery records
- 10 key features for modeling
- Realistic patterns and distributions
- Business context for analysis

Jedha

Why This Assessment Approach?

Realistic & Practical

- Real-world scenario with time constraints
- End-to-end workflow from data creation to analysis
- Interconnected projects reflecting actual workplace dynamics

Comprehensive Skill Evaluation

- Tests technical abilities across multiple domains
- Evaluates decision-making under pressure
- Assesses code quality and documentation practices

Jedha

Expected Outcomes

For Data Engineers

- Functional ETL pipeline
- Clean, documented dataset
- Code demonstrating DE best practices
- Evidence of certification competencies

For Data Scientists

- Insightful data analysis
- Working predictive model
- Clear visualization of findings
- Actionable business recommendations

Jedha

Assessment Criteria

Data Engineering

- Code quality and structure
- Data pipeline design
- Transformation logic
- Error handling
- Documentation quality
- Data validation approach

Data Science

- Analysis depth and relevance
- Visualization quality
- Model performance
- Feature selection approach
- Clarity of insights
- Business value of recommendations

Learning Outcomes

Data Engineers will demonstrate ability to:

- Design and implement data pipelines under time constraints
- Transform and integrate data from multiple sources
- Apply data quality principles
- Document technical implementations

Data Scientists will demonstrate ability to:

- Extract insights from complex datasets
- Develop predictive models for business problems
- Communicate findings effectively
- Deliver actionable recommendations

Jedha

Get Ready!

You have 2 hours to complete your project!

- Data Engineers: Create the dataset that Data Scientists need
- Data Scientists: Analyze the data and build your predictive model
- Both: Document your approach and results

Remember:

Focus on demonstrating the competencies in your certification blocks!

Jedha

Questions?

Resources:

- Starter code is provided
- Documentation references available
- Mentors on standby for critical issues

Good luck!