Data Literacy Body of Knowledge

eLearningCurve

1. Data and Databases

1.1 Data Fundamentals

1.1.1 Kinds of Data

- 1.1.1.1 Data vs. Information
- 1.1.1.2 Raw Data and Refined Data
- 1.1.1.3 Atomic, Summary, and Aggregated
- 1.1.1.4 Event data and Reference Data
- 1.1.1.5 Transactional, Operational and Analytical Data

1.1.2 Provenance of Data

- 1.1.2.1 Internal – Enterprise, Departmental, Local, Individual, etc.
- 1.1.2.2 External – Partner, Commercial, Open, Public, etc.

1.1.3 Data Organization

- 1.1.3.1 Structured and Unstructured
- 1.1.3.2 Semi-Structured and Multi-Structured

1.1.4 Data Contents

- 1.1.4.1 Data Types
- 1.1.4.2 Business Data vs. Metadata
- 1.1.4.3 Variables – Quantitative, Qualitative, and Categorical
- 1.1.4.4 Variables – Continuous and Discrete

1.2 Database Fundamentals

1.2.1 Common File Formats

- 1.2.1.1 Fixed Format
- 1.2.1.2 Delimited – comma delimited (CSV), pipe delimited, tab delimited (TSV), etc.
- 1.2.1.3 Self-Describing – XML, JSON

1.2.2 Spreadsheets

- 1.2.2.1 Rows and Columns
- 1.2.2.2 Formulas and Macros
- 1.2.2.3 Pivot Tables

© 2021 by eLearningCurve LLC. All rights reserved. Reproduction in whole or part prohibited except by written permission.

1.2.2.4 Worksheets and Workbooks

1.2.3 Relational Databases

1.2.3.1 Tables, Rows, and Columns
1.2.3.2 Keys, Foreign Keys, and Relationships
1.2.3.3 Structured Query Language (SQL)

1.2.4 Multi-Dimensional Databases

1.2.4.1 Fact Tables and Facts
1.2.4.2 Dimensions and Dimension Attributes
1.2.4.3 Dimension Hierarchy
1.2.4.4 OLAP Query and Analysis

1.2.5 NoSQL Databases

1.2.5.1 Key-Value Stores
1.2.5.2 Document Stores
1.2.5.3 Wide-Column Stores
1.2.5.4 Graph Databases

2. **Data Knowledge and Data Governance**

2.1 Managing Data Knowledge

2.1.1 Metadata Management

2.1.1.1 What is Metadata?
2.1.1.2 Kinds of Metadata
2.1.1.3 Metadata Processes
2.1.1.4 Roles of Metadata
2.1.1.5 Metadata Management Tools

2.1.2 Data Cataloging

2.1.2.1 What is a Data Catalog?
2.1.2.2 Data Catalog Metadata
2.1.2.3 Data Catalog Functions
2.1.2.4 Data Catalog Benefits

2.2 Data Governance

© 2021 by eLearningCurve LLC. All rights reserved. Reproduction in whole or part prohibited except by written permission.

2.2.1 Data Governance Basics

2.2.2.1 What is Data Governance?
2.2.2.3 Why is Data Governance Needed?
2.2.2.4 People, Processes, and Technology in Data Governance

2.2.2 Data Governance Goals and Purpose

2.2.2.1 Data Protection – Security, Privacy, and Compliance
2.2.2.2 Data Utility – Quality, Integration, and Metadata
2.2.2.3 Data Value – Impact, Risk, and Retention

2.2.3 Data Governance Processes and Practices

2.2.3.1 Data Policy Management
2.2.3.2 Data Curation and Cataloging
2.2.3.4 Data Quality Management
2.2.3.5 Data Stewardship
2.2.3.6 Data Ethics

2.2.4 Data Consumer Responsibilities

2.2.4.1 Regulatory Compliance
2.2.4.2 Policy Compliance
2.2.4.3 Data Constraints and Appropriate Use
2.2.4.4 Evaluating and Understanding Data
2.2.4.5 Data Sharing
2.2.4.6 Knowledge Sharing

3. **Data Resource Management**

3.1 Data Resource Consolidation

3.1.1 Enterprise Data Resource Basics

3.1.1.1 What is the Enterprise Data Resource?
3.1.1.2 Why Enterprise Data Resource?
3.1.1.3 People, Processes, and Technology

3.1.2 Data Consolidation

3.1.2.1 What is Data Consolidation?

© 2021 by eLearningCurve LLC. All rights reserved. Reproduction in whole or part prohibited except by written permission.

3.1.2.2 Why Consolidate Data

3.2 Managing the Data Resource

3.2.1 Data Resource Management Architectures

3.2.1.1 Data Warehousing
3.2.1.2 Data Lakes
3.2.1.3 Master Data Management (MDM)

3.2.2 Data Resource Management Processes

3.2.2.1 Data Integration
3.2.2.2 Data Engineering
3.2.2.3 Data Preparation

3.2.3 Sharing the Data Resource

3.2.3.1 Shared Resources
3.2.3.2 Shared Knowledge
3.2.3.3 Shared Files
3.2.3.4 Shared Reporting and Analysis

3.3 Using the Data Resource

3.3.1 Business Intelligence (BI)

3.3.1.1 What and Why of BI
3.3.1.2 Query and Reporting
3.3.1.3 OLAP
3.3.1.4 Monitoring and Alerts

3.3.2 Performance Management

3.3.2.1 What and Why of Performance Management
3.3.2.2 Performance Dashboards and Scorecards

3.3.3 Business Analytics

3.3.3.1The What and Why of Business Analytics
3.3.3.2 Kinds of Analytics – Descriptive, Diagnostic, Predictive, Prescriptive

3.3.4 Advanced Analytics and Data Science

© 2021 by eLearningCurve LLC. All rights reserved. Reproduction in whole or part prohibited except by written permission.

3.3.4.1 Data Mining
3.3.4.2 Prediction, Prescription, and Automation
3.3.4.3 Artificial Intelligence and Machine Learning

4. Data Provisioning

4.1 Finding and Evaluating Data

4.1.1 Project Framing

4.1.1.1 Analysis Goals – Explore, Learn, Answer Questions
4.1.1.2 Problem Framing – Stating the Analysis Problem

4.1.2 Defining Requirements

4.1.2.1 Business Requirements – What do you need to do?
4.1.2.2 Information Requirements – What do you need to know?
4.1.2.3 Data Requirements – What data do you need?

4.1.3 Searching for Data

4.1.3.1 Crowdsourcing
4.1.3.2 Data Catalog
4.1.3.3 Data Marketplace

4.1.4 Evaluating Data

4.1.4.1 Content
4.1.4.2 Timeliness – Freshness, History & Time Spans, etc.
4.1.4.3 Quality
4.1.4.4 Metadata and Lineage
4.1.4.5 Support – Stewards, SMEs, etc.
4.1.4.6 Trust

4.2 Data Preparation

4.2.1 Exploring and Profiling Data

4.2.1.1 Data Understanding – Content, Structure, Data Types, etc.
4.2.1.2 Data Discovery – Patterns, Relationships, Anomalies
4.2.1.3 Data Profiles

4.2.2 Transforming Data

© 2021 by eLearningCurve LLC. All rights reserved. Reproduction in whole or part prohibited except by written permission.

© 2021 by eLearningCurve LLC. All rights reserved. Reproduction in whole or part prohibited except by written permission.

5.1.3.2 Correlation
5.1.3.3 Variance
5.1.3.4 Co-Variance

5.1.4 Time Series Analysis

5.1.4.1 Moving Average Techniques
5.1.4.2 Univariate Time Series Models
5.1.4.3 Multivariate Time Series Models

5.1.5 Data Analysis Technologies

5.1.5.1 Spreadsheet Analysis
5.1.5.2 Self-Service Analysis Tools
5.1.5.3 Advanced Analytics and Data Science Tools

5.2 Data Visualization

5.2.1 Data Visualization Functions

5.2.1.1 Comparisons and Distributions
5.2.1.2 Relationships, Proportions, and Compositions
5.2.1.3 Patterns, Trends, and Time Series

5.2.2 Charts and Graphs

5.2.2.1 Types of Charts and Graphs – Line, Bar, Column, Scatter, Map, etc.
5.2.2.2 Components of Charts and Graphs – Axes, Scales, Points, Lines, Shapes, Title,
5.2.2.3 Legend, Annotations
5.2.2.4 Visual Composition – Colors, Fonts, Aspect Ratio, etc.

5.2.3 Reading Data Visualizations

5.2.3.1 Context – Title, Axes, Scales, Legend
5.2.3.2 Data – Variables, Quantities, Units
5.2.3.3 Meaning – Patterns, Trends, Inference

5.2.4 Creating Data Visualizations

5.2.4.1 Choosing the Visualization Type
5.2.4.2 Defining the Axes
5.2.4.3 Defining the Scales – Linear and Log, Zero and Non-Zero Base
5.2.4.4 Units of Measure
5.2.4.5 Title, Legend, Annotations

© 2021 by eLearningCurve LLC. All rights reserved. Reproduction in whole or part prohibited except by written permission.

5.2.4.6 User Interaction – Navigate, Filter, Drill, etc.
5.2.4.7 Content, Composition, and Layout
5.2.4.8 Avoiding the Pitfalls

5.3 Analysis to Action

5.3.1 Data Storytelling

5.3.1.1 Interpreting the Analysis
5.3.1.2 Conclusions and Recommendations
5.3.1.3 Crafting the Narrative
5.3.1.4 Telling the Story

5.3.2 Acting on Information and Analysis

5.3.2.1 Informed Decision Making
5.3.2.2 Data-Driven Recommendations
5.3.2.3 Data-Driven Automation
5.3.2.4 Data-Driven Innovation

5.3.3 Verification and Adaptation

5.3.3.1 Business Goals
5.3.3.2 Indicators, Metrics, and Measures
5.3.3.3 Feedback Loop

© 2021 by eLearningCurve LLC. All rights reserved. Reproduction in whole or part prohibited except by written permission.