|  |  |  |
| --- | --- | --- |
| Dashboard / My courses / Tsel\_DSA / Topic 27 ; Big data pipeline development II  / Quiz Big data pipeline development II | | |
|  | **Started on** Friday, 24 January 2020, 6;09 PM  **State** Finished  **Completed on** Friday, 24 January 2020, 6;10 PM  **Time taken** 41 secs  **Marks** 5.00/5.00  **Grade 10.00** out of 10.00 (**100**%)  Question **1**  Complete  Mark 1.00 out of 1.00  Whats --hive-import argument really do?  Select one;   1. Import tables into hive with default delimiters if none are set 2. Import tables into hive with default delimiters 3. Import tables into hive with specific delimiters   The correct answer is; Import tables into hive with default delimiters if none are set |  |



Question **2** Complete

Mark 1.00 out of 1.00



sqoop import-all-tables -m 1 --connect jdbc;postgresql;//host;port /postgres --username iykra\_trainee --password passwordiykra --warehouse-dir user/hive/app\_loan2 schema training

Whats the database are imported?

Select one;

1. app\_loan2
2. postgres
3. PostgresSQL

The correct answer is; postgres

Question **S** Complete

Mark 1.00 out of 1.00



sqoop import -m 1 --connect jdbc;postgresql;//host;port /postgres --username iykra\_trainee -- password passwordiykra --target-dir user/hive/app\_loan2

Whats the output of above syntax?

Select one;

1. Error
2. Table in HIVE
3. File that include all values in specific table
4. Empty file

The correct answer is; Error

|  |  |  |
| --- | --- | --- |
|  | **Quiz Big data pipeline development I**  Jump to...  **Quiz Big data pipeline development III €** |  |
|  | | |
| Copyright © 2019 IYKRA | | |



Question **4** Complete

Mark 1.00 out of 1.00



sqoop import-all-tables -m 8 --connect jdbc;postgresql;//host;port /postgres --username iykra\_trainee --password passwordiykra --warehouse-dir user/hive/app\_loan2 schema training

How much map task are used from above syntax?

Select one;

1. 1
2. 8
3. 3
4. 4

Dashboard / My courses / Tsel\_DSA / Topic 26 ; Big data pipeline development I

/ Quiz Big data pipeline development I

**Started on** Friday, 24 January 2020, 6;14 PM



**State** Finished

**Completed on** Friday, 24 January 2020, 6;14 PM

**Time taken** 51 secs

**Marks** 5.00/5.00

**Grade 10.00** out of 10.00 (**100**%)

The correct answer is; 8

Question **1**

Complete

Mark 1.00 out of 1.00

sqoop import-all-tables -m 1 --connect jdbc;postgresql;//host;port /postgres --username iykra\_trainee --password passwordiykra --warehouse-dir user/hive/app\_loan2 schema training

Whats the directory of the output file from above syntax?

Select one;

1. host;port/postgres
2. user/hive/app\_loan2
3. //host;port

The correct answer is; user/hive/app\_loan2

Question **5** Complete

Mark 1.00 out of 1.00

The correct answer is; 1



sqoop import- -m 8 --connect jdbc;postgresql;//host;port /postgres --username iykra\_trainee -- password passwordiykra --target -dir user/hive/app\_loan2 -table application\_loan

If chosen schema have 4 tables on postgreSQL, how much the file part will be on the output directory?

Select one;

1. 4
2. 32
3. 12
4. 1

Question **2** Complete

Mark 1.00 out of 1.00



Why we need disable SElinux before we install cloudera?

Select one;

1. We need to disable the SElinux to enhance node performance while installing Hadoop
2. SElinux will thread cloudera as virus
3. No need to do that

The correct answer is; SElinux will thread cloudera as virus

Question **S** Complete

Mark 1.00 out of 1.00



What is the di"rence between external and internal table?

Select one;

1. The data from internal table was not disappear if we drop the table, while the external table was
2. Nothing, exactly the same
3. The data from external table was not if we drop the table, while the internal table was

The correct answer is; The data from external table was not disappear if we drop the table, while the internal table was

Question **4** Complete



Mark 1.00 out of 1.00

Define sqoop?

Select one;

1. Application for transfering bulk data between hadoop and non hadoop datastores
2. Application for transfering bulk data in hadoop ecosystem
3. Application for distributing file in hdfs

The correct answer is; Application for transfering bulk data between hadoop and non hadoop datastores

Question **5** Complete

Mark 1.00 out of 1.00

Whats the di"erent between sqoop import and sqoop import-all-table?

Select one;

1. Transfering all table value from specific database non hadoop
2. Transfering all table from specific database on hadoop
3. Transfering all databases on hadoop

The correct answer is; Transfering all table value from specific database non hadoop

# Quiz Data visualization for analysis I

Jump to...

# Quiz Big data pipeline development II €



Dashboard / My courses / Tsel\_DSA / Topic 28 ; Big data pipeline development III

/ Quiz Big data pipeline development III

**Started on** Friday, 24 January 2020, 6;04 PM

**State** Finished

**Completed on** Friday, 24 January 2020, 6;04 PM

**Time taken** 50 secs

**Marks** 5.00/5.00

**Grade 10.00** out of 10.00 (**100**%)

Question **1**

Complete

Mark 1.00 out of 1.00

The following entity is included in tree, exclude;

Select one;

1. Branch
2. Probability
3. Node
4. Leaf

The correct answer is; Probability

Question **2** Complete

Mark 1.00 out of 1.00



Whats the meaning of 'pure dataset' in decision tree?

Select one;

1. Perfectly split based on "rules"
2. All answer are true
3. Clean data
4. Dynamic dataset

The correct answer is; Clean data

Question **S** Complete

Mark 1.00 out of 1.00



The following method is used to Attribute Selection Measures

Select one;

1. All answer are true
2. P value
3. Confidence Interval
4. Entropy

The correct answer is; Entropy

Question **4** Complete



Mark 1.00 out of 1.00

What is the parent node means?

Select one;

1. The feaeture variable that we have to split
2. The train dataset
3. All answer are true
4. The target variable that we want to predict

The correct answer is; The target variable that we want to predict

Question **5** Complete

Mark 1.00 out of 1.00

The advantage for using Decision Tree;

Select one;

1. Need to use normalization
2. Can model non linear relationship
3. Black box, hard to interpret model

The correct answer is; Can model non linear relationship

# Quiz Big data pipeline development II

Jump to...

# Quiz Business Intelligence I €

Copyright © 2019 IYKRA

Copyright © 2019 IYKRA



Dashboard / My courses / Tsel\_DSA / Topic 29 ; Business Intelligence I

/ Quiz Business Intelligence I

**Started on** Friday, 24 January 2020, 5;57 PM

**State** Finished

**Completed on** Friday, 24 January 2020, 6;00 PM

**Time taken** 2 mins 23 secs

**Marks** 5.00/5.00

**Grade 10.00** out of 10.00 (**100**%)

Question **1**

Complete

Mark 1.00 out of 1.00

Contoh hasil case study predictive analytics dibawah ini adalah..

Select one;

1. Jumlah Penduduk Indonesia menurut Provinsi Tahun 2010
2. Persebaran Order Ojek Online di Jakarta Tahun 2018
3. Credit Risk Scoring Bank X untuk Produk Unsecured Loan
4. Jumlah Penggunaan Obat Generik di Surabaya Tahun 2012

The correct answer is; Credit Risk Scoring Bank X untuk Produk Unsecured Loan

Question **2** Complete

Mark 1.00 out of 1.00



CRISP-DM adalah

Select one;

1. Cross Section Process for Data Mining
2. Cross Industry Standard Process for Data Mining
3. Cross Tabulation Standard Process for Data Mining
4. Cross Industry Sectional Process for Data Mining

The correct answer is; Cross Industry Standard Process for Data Mining

Question **S** Complete

Mark 1.00 out of 1.00



Step pertama dalam Business Analytics Workflow adalah

Select one;

1. Data Presentation
2. Frame The Problem
3. Deliver The Results
4. Machine Learning

The correct answer is; Frame The Problem

Question **4** Complete



Mark 1.00 out of 1.00

Berapa jumlah steps yang ada di dalam CRISP-DM ?

Select one;

1. 5
2. 4
3. 6
4. 3

The correct answer is; 6

Question **5** Complete

Mark 1.00 out of 1.00

"Ada sebuah studi kasus tentang forecasting demand dari suatu barang di perusahaan e- commerce," ini termasuk tipe analytics..

Select one;

1. predictive
2. descriptive
3. initiative
4. explorative

The correct answer is; predictive

# Quiz Big data pipeline development III

Jump to...

# Quiz Scale up data science I €

Dashboard / My courses / Tsel\_DSA / Topic 33 ; Business Intelligence II

/ Quiz Business Intelligence II

**Started on** Friday, 24 January 2020, 5;30 PM



**State** Finished

**Completed on** Friday, 24 January 2020, 5;33 PM

**Time taken** 2 mins 41 secs

**Marks** 5.00/5.00

**Grade 10.00** out of 10.00 (**100**%)

Question **1**

Complete

Mark 1.00 out of 1.00

Jenis data apa yang dapat digunakan untuk membuat visualisasi?

Select one;

1. Semua benar
2. Data kualitatif dan kuantitatif
3. Data diskrit dan kontinyu
4. Data internal dan eksternal

The correct answer is; Semua benar

Question **2** Complete

Mark 1.00 out of 1.00



Contoh dari data Internal adalah ;

Select one;

1. Jumlah karyawan
2. Daya beli masyarakat
3. Tingkat inflasi
4. Jumlah penduduk kota

The correct answer is; Jumlah karyawan

Question **S** Complete

Mark 1.00 out of 1.00



Contoh data berkala (time series data) adalah

Select one;

1. Semua benar
2. Laporan keuangan penjualan tahun 2016-2018
3. Laporan keuangan perusahaan posisi 31 Desember 2016 dan 31 Desember 2017
4. Perkembangan nilai tukar Rupiah terhadap US Dollar

The correct answer is; Semua benar

Question **4** Complete



Mark 1.00 out of 1.00

Software-software dibawah ini dapat digunakan untuk membuat visualisasi data, kecuali ...

Select one;

1. Microsoft Excel
2. Microsoft Power BI
3. Microsoft Word
4. Microsoft Power Point

The correct answer is; Microsoft Word

Question **5** Complete

Mark 1.00 out of 1.00

Tujuan utama dari visualisasi data adalah ;.

Select one;

1. Semua benar
2. membantu pengguna dalam menganalisa dan penalaran tentang data
3. membuat data yang kompleks menjadi mudah dipahami dan berguna
4. untuk mengkomunikasikan informasi secara jelas dan efisien kepada pengguna lewat grafik

The correct answer is; Semua benar

# Quiz Scale up data science III

Jump to...

# Quiz Business Intelligence III €

Copyright © 2019 IYKRA

Dashboard / My courses / Tsel\_DSA / Topic 34 ; Business Intelligence III

/ Quiz Business Intelligence III

**Started on** Friday, 24 January 2020, 5;24 PM



**State** Finished

**Completed on** Friday, 24 January 2020, 5;28 PM

**Time taken** 3 mins 13 secs

**Marks** 4.00/5.00

**Grade 8.00** out of 10.00 (**80**%)

Question **1**

Complete

Mark 1.00 out of 1.00

Common definition of big data?

Select one;

1. Data that is too large or too complex to be managed using traditional data processing, analysis, and storage techniques
2. Set of the data
3. Data that is simple to be managed using traditional data processing, analysis, and storage techniques
4. Data that is easy to be managed using traditional data processing, analysis, and storage techniques

The correct answer is; Data that is too large or too complex to be managed using traditional data processing, analysis, and storage techniques

Question **2** Complete

Mark 1.00 out of 1.00



What's the function of Map/Reduce?

Select one;

1. Provides parallel processing of data in HDFS.
2. Data warehousing tools.
3. NoSQL database.
4. To analyze data sets.

The correct answer is; Provides parallel processing of data in HDFS.

Question **S** Complete

Mark 1.00 out of 1.00



Whats volume of data really say in terms of Big Data?

Select one;

1. How data correction works in Big Data.
2. All answers are true
3. Volume of data.
4. Not only the volume of data but also the correction of the data.

The correct answer is; Volume of data.

Question **4** Complete



Mark 1.00 out of 1.00

Di"erent types and forms of data dikenal dengan istilah?

Select one;

1. Volume
2. Velocity
3. Variety
4. Veracity

The correct answer is; Variety

Question **5** Complete

Mark 0.00 out of 1.00

Data Scientist menangani Data, Information, Knowledge dan Understanding yang bersifat?

Select one;

1. Unknown Unknowns
2. Known Knowns
3. Known Unknowns
4. Known Unknowns dan Unknown Unknowns

The correct answer is; Known Unknowns dan Unknown Unknowns

# Quiz Business Intelligence II

Jump to...

# Quiz Data Science Delivery I €

Copyright © 2019 IYKRA

Dashboard / My courses / Tsel\_DSA / Topic 35 ; Data Science Delivery I

/ Quiz Data Science Delivery I

**Started on** Friday, 24 January 2020, 5;18 PM



**State** Finished

**Completed on** Friday, 24 January 2020, 5;22 PM

**Time taken** 4 mins 52 secs

**Marks** 4.00/5.00

**Grade 8.00** out of 10.00 (**80**%)

Question **1**

Complete

Mark 1.00 out of 1.00

metodologi project yang digunakan untuk sebuah data science project adalah

Select one;

1. waterfall atau MVP
2. agile atau watermark
3. agile atau MVP
4. agile atau waterfall

The correct answer is; agile atau waterfall

Question **2** Complete

Mark 1.00 out of 1.00



PMML adalah

Select one;

1. Predictive Model Markup Language
2. Predictive Model Management Language
3. Prescriptive Model Management Language
4. Prescriptive Model Markup Language

The correct answer is; Predictive Model Markup Language

Question **S** Complete

Mark 0.00 out of 1.00



Jika kita melakukan training model di Jupyter Notebook, lalu akan melakukan deployment di Application dengan Programming Language Java, maka disarankan kita mengeluarkan output model dalam format

Select one;

1. HTML
2. Pickle
3. JSON
4. PMML

The correct answer is; PMML

Question **4** Complete



Mark 1.00 out of 1.00

Berikut ini adalah contoh output dari sebuah data science project kecuali

Select one;

1. Aplikasi
2. Archive File
3. Dashboard
4. Model

The correct answer is; Archive File

Question **5** Complete

Mark 1.00 out of 1.00

Seseorang yang bertugas mengatur workload dari sebuah Data Science Team disebut

Select one;

1. Project Manager
2. Data Engineer
3. Product Manager
4. Project Assistant

The correct answer is; Project Manager

# Quiz Business Intelligence III

Jump to...

# Quiz Data Science Delivery II €

Copyright © 2019 IYKRA



Dashboard / My courses / Tsel\_DSA / Topic 36 ; Data Science Delivery II

/ Quiz Data Science Delivery II

**Started on** Friday, 24 January 2020, 4;58 PM

**State** Finished

**Completed on** Friday, 24 January 2020, 5;16 PM

**Time taken** 18 mins 9 secs

**Marks** 4.00/5.00

**Grade 8.00** out of 10.00 (**80**%)

Question **1**

Complete

Mark 1.00 out of 1.00

"process of using domain knowledge of the data to create features that make machine learning algorithms work" adalah definisi dari

Select one;

1. feature system
2. feature filtering
3. feature engineering
4. feature extraction

The correct answer is; feature engineering

Question **2** Complete

Mark 1.00 out of 1.00



Contoh visualisasi data yang penting dalam sebuah Data Science Project di bawah ini, kecuali

Select one;

1. Lorentz Curve
2. Distribution Plot
3. Correlation Matrix
4. ROC Curve

The correct answer is; Lorentz Curve

Question **S** Complete

Mark 1.00 out of 1.00



Dalam proses model deployment, tujuan dari pickling model adalah

Select one;

1. Meminimalisir size dari suatu model
2. Memindahkan model ke suatu aplikasi java
3. Membuat model menjadi readable
4. Serialisasi python object agar dapat di load ke python program

The correct answer is; Serialisasi python object agar dapat di load ke python program

Question **4** Complete



Mark 0.00 out of 1.00

Berikut adalah syntax yang benar untuk write model ke dalam bentuk pickle

Select one;

1. pd.write\_models('models.pkl')
2. pickle.dump(Model, open('Models.pkl', 'wb'))
3. pickle.loads(f.read())
4. joblib.load('Models.pkl')

The correct answer is; pickle.dump(Model, open('Models.pkl', 'wb'))

Question **5** Complete

Mark 1.00 out of 1.00

Berikut adalah cara penghitungan F1 Score

Select one;

1. 2\*(Recall \* Precision) / (Recall + Precision)
2. TP+TN/(TP+FP+FN+TN)
3. TP/(TP+FP)
4. (TP+TN)/(TP\*TN)

The correct answer is; 2\*(Recall \* Precision) / (Recall + Precision)

# Quiz Data Science Delivery I

Jump to...

# Quiz Data Science Delivery III €

Copyright © 2019 IYKRA



|  |  |  |
| --- | --- | --- |
| Dashboard / My courses / Tsel\_DSA / Topic 38 ; Soft Skills for data scientist II  / Quiz Data Science Delivery III | | |
|  | **Started on** Wednesday, 22 January 2020, 10;36 AM  **State** Finished  **Completed on** Wednesday, 22 January 2020, 10;47 AM  **Time taken** 11 mins 7 secs  Question **1**  Complete  Marked out of 1.00  Penentuan sebuah "Business Success Criteria" termasuk dalam scope metodologi CRISP-DM pada fase  Select one;   1. Business Understanding 2. Data Understanding 3. Evaluation 4. Business Solution |  |

Question **2** Complete

Marked out of 1.00



Langkah yang tidak kalah penting dalam workflow sebuah Data Science Project yaitu sebuah mekanisme tracking serta pemeriksaan berkala dari model machine learning yang sudah di deploy disebut.... Dan termasuk dalam scope metodologi CRISP-DM pada fase....

Select one;

1. Evaluate Results - Deployment
2. Plan Monitoring and Maintenance - Deployment
3. Evaluate Results - Evaluation
4. Plan Monitoring and Maintenance - Evaluation

Question **S** Complete

Marked out of 1.00



Berikut adalah syntax yang benar untuk write model ke dalam bentuk pickle

Select one;

1. pd.write\_models('models.pkl')
2. joblib.load('Models.pkl')
3. pickle.loads(f.read())
4. pickle.dump(Model, open('Models.pkl', 'wb'))

Question **4** Complete



Marked out of 1.00

Generate nilai COUNT, MIN, MAX dari sebuah variabel adalah salah satu contoh Feature Engineering

Select one;

1. Domain Knowledge
2. Intermediate
3. Advanced
4. Basic

Question **5** Complete

Marked out of 1.00

Berikut adalah fakta yang benar mengenai project management Agile vs Waterfall

Select one;

1. Agile mempunyai peluang resiko lebih rendah, Waterfall lebih tinggi
2. Waterfall adalah change-driven approach
3. Agile adalah plan-driven approach
4. Agile mempunyai peluang resiko project lebih tinggi, Waterfall lebih rendah

# Quiz Data Science Delivery III

Jump to...

# Quiz Soft Skills for data scientist III €

Copyright © 2019 IYKRA



Dashboard / My courses / Tsel\_DSA / Topic 24 ; Data solution & architecture II

/ Quiz Data solution & architecture II

**Started on** Friday, 24 January 2020, 6;21 PM

**State** Finished

**Completed on** Friday, 24 January 2020, 6;22 PM

**Time taken** 38 secs

**Marks** 5.00/5.00

**Grade 10.00** out of 10.00 (**100**%)

Question **1**

Complete

Mark 1.00 out of 1.00

Berikut ini yang tidak termasuk komponen penting untuk dipahami untuk menyelesaikan business terkait dengan data warehouse modernization

Select one;

1. Flow
2. Operating System
3. Data Tech Component
4. Architecture

The correct answer is; Operating System

Question **2** Complete

Mark 1.00 out of 1.00



Data flow terdiri atas

Select one;

1. OLAP & BI
2. OLAP & OLDP
3. OLAP & OLTP
4. OLTP & OLDP

The correct answer is; OLAP & OLTP

Question **S** Complete

Mark 1.00 out of 1.00



Operational Database umumnya menggunakan flow..

Select one;

1. OLL
2. OLAP
3. OLDP
4. OLTP

The correct answer is; OLTP

Question **4** Complete



Mark 1.00 out of 1.00

Apache Kafka adalah aplikasi yang mendukung

Select one;

1. Descriptive analytics
2. Real-time streaming process
3. Data point expansion
4. Batch process

The correct answer is; Real-time streaming process

Question **5** Complete

Mark 1.00 out of 1.00

Data delivery guarantee dengan konsep "minimal satu kali" pengiriman adalah ;

Select one;

1. At least once
2. Best Practice
3. Best e"ort
4. Exaclty once

The correct answer is; At least once

# Quiz Machine Learning VI

Jump to...

# Quiz Data visualization for analysis I €

Copyright © 2019 IYKRA

|  |  |  |
| --- | --- | --- |
| Dashboard / My courses / Tsel\_DSA / Topic 20 ; Data Solution and Architecture I  / Quiz Data Solution and Architecture I | | |
|  | **Started on** Friday, 24 January 2020, 6;40 PM  **State** Finished  **Completed on** Friday, 24 January 2020, 6;41 PM  **Time taken** 49 secs  **Marks** 5.00/5.00  **Grade 10.00** out of 10.00 (**100**%) |  |

Question **1** Complete

Mark 1.00 out of 1.00



Which one define the relationship between data governance and data management? Select one;

a.

Both are related and inseparable

b.

Both are unrelated subjects

c.

Both are related but has to implemented exclusively

d.

None of them are right

The correct answer is;

Both are related and inseparable

Question **2** Complete

Mark 1.00 out of 1.00



Which one IS NOT included in elements of Data Governance?

Select one;

a.

None of the above

b.

Data quality

c.

Data strategy

d.

Data risk management

The correct answer is; None of the above

Question **S** Complete

Mark 1.00 out of 1.00



What is the goal of Data Warehousing?

Select one;

a.

Achieve company´s endgame

b.

Store specific data for each business functions

c.

Store all of the company´s data from across services

d.

Generating business insights

The correct answer is;

Store all of the company´s data from across services

Question **4** Complete

Mark 1.00 out of 1.00



Which one IS NOT included in BI Architecture Framework

Select one;

a.

Integrate

b. Create

c.

Distribute

d. Store

The correct answer is; Create

Question **5** Complete



Mark 1.00 out of 1.00

What is the goal of Data Pipeline?

Select one;

a.

Create journey from raw data to the data warehouse

b.

Extract data from various sources

c.

Standardized data from various sources

d.

Load data to data warehouse

The correct answer is;

Create journey from raw data to the data warehouse

**Mid-Test Statistika** Jump to...

**Quiz Machine Learning IV €**

Copyright © 2019 IYKRA

Dashboard / My courses / Tsel\_DSA / Topic 25 ; Data visualization for analysis I

/ Quiz Data visualization for analysis I

**Started on** Friday, 24 January 2020, 6;18 PM

**State** Finished

**Completed on** Friday, 24 January 2020, 6;18 PM

**Time taken** 40 secs

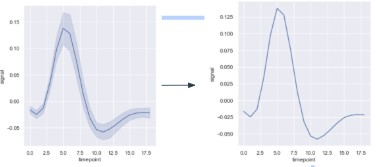
**Marks** 5.00/5.00

**Grade 10.00** out of 10.00 (**100**%)

Question **1**

Complete

Mark 1.00 out of 1.00



What command to turn o" the shades;

Select one;

1. **ci = None**
2. **shades = None**
3. **conf.int = None**
4. **ci = False**
5. **shades = False**

The correct answer is; **ci = None**

Question **2**

Complete

Mark 1.00 out of 1.00



Which statement is correct;

Select one;

1. **Seaborn package requires Matplotlib package to run**
2. **Matplotlib package requires Seaborn package to run**

The correct answer is; **Seaborn package requires Matplotlib package to run**

Question **S**

Complete

Mark 1.00 out of 1.00



Which statement is wrong;

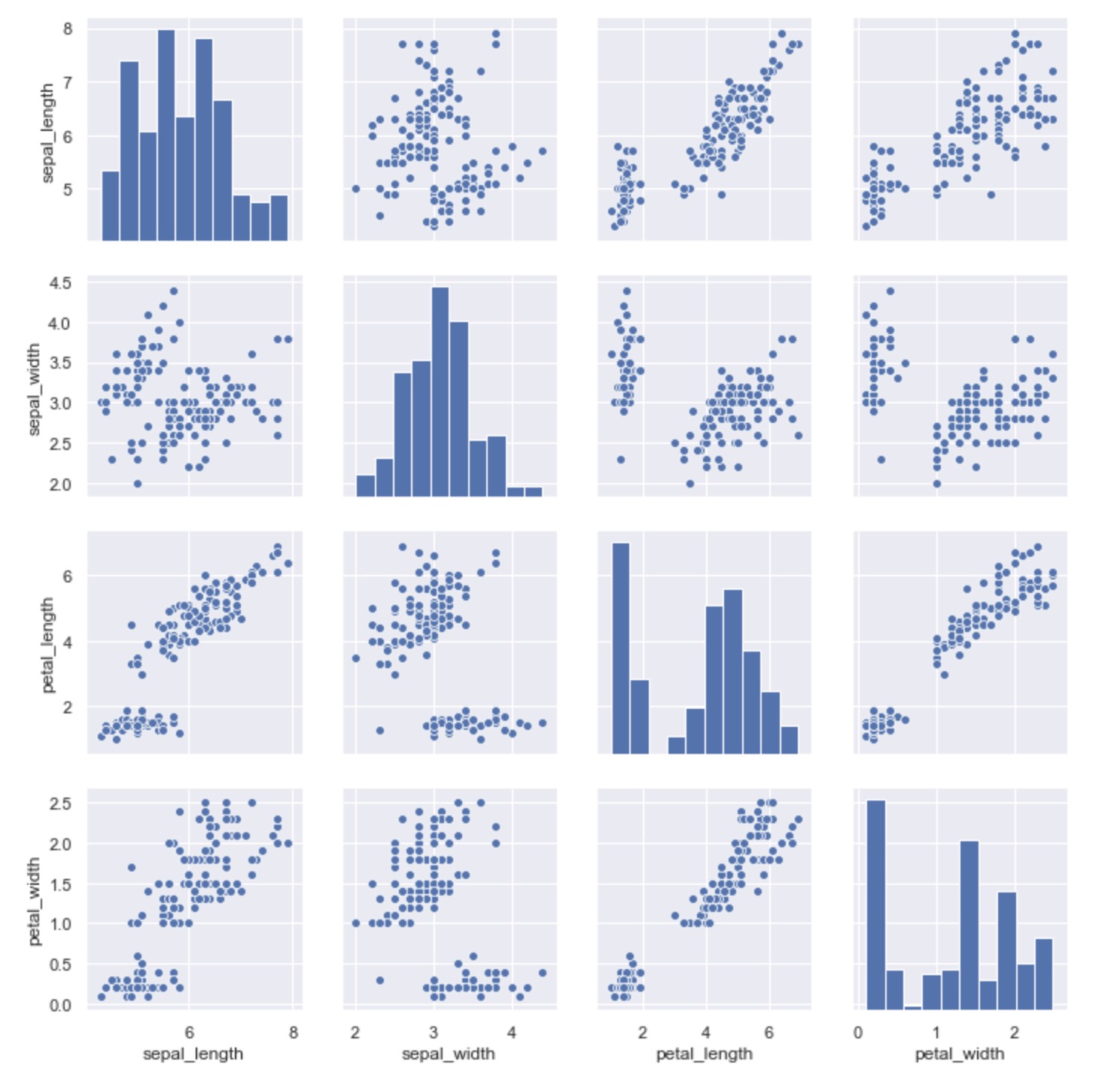
Select one;

1. **Joint plot can show the distribution relationship up to S distributions**
2. **Box plot helps the user to spot outliers**
3. **Violin plot allows the user to identify bi-modal distribution**

The correct answer is; **Joint plot can show the distribution relationship up to S distributions**

Question **4** Complete

Mark 1.00 out of 1.00



What Seaborn function to produce plot like this?

Select one;

1. sns.facet\_wrap()
2. sns.jointplot()
3. sns.pairplot()

The correct answer is; sns.pairplot()

Question **5** Complete



Mark 1.00 out of 1.00

What Seaborn command to color-code di"erent groups?

Select one;

1. hue
2. col
3. kind

The correct answer is; hue

# Quiz Data solution & architecture II

Jump to...

# Quiz Big data pipeline development I €

Copyright © 2019 IYKRA



Dashboard / My courses / Tsel\_DSA / Topic 12 ; Data Wrangling I I

/ Quiz Intro to Data Engineering and ETL Fundamentals using Python I

**Started on** Friday, 24 January 2020, 7;17 PM

**State** Finished

**Completed on** Friday, 24 January 2020, 7;19 PM

**Time taken** 2 mins 14 secs

**Marks** 4.00/5.00

**Grade 8.00** out of 10.00 (**80**%)

Question **1**

Complete

Mark 1.00 out of 1.00

Which one is tool for workflow orchestration?

Select one;

1. **Apache Spark**
2. **Apache Airflow**
3. **Apache Hive**

The correct answer is; **Apache Airflow**

Question **2**

Complete

Mark 0.00 out of 1.00



Which one of the following is Cron expression for every Sunday at 7.15 PM?

Select one;

a. **0 \* \* 7 15**

b. **15 19 \* \* 0**

c. **15 7 \* \* 0**

The correct answer is; **15 19 \* \* 0**

Question **S**

Complete

Mark 1.00 out of 1.00



How to delete column in Spark DataFrame?

Select one;

1. **df.delete("awesome\_column")**
2. **df.drop("awesome\_column")**
3. **df.remove("awesome\_column")**

The correct answer is; **df.drop("awesome\_column")**

Question **4**

Complete

Mark 1.00 out of 1.00



What is the responsibilities that suits data engineer the most?

Select one;

1. **Come up with a database schema for an application.**
2. **Set up scheduled ingestion of data from the application databases to an analytical database**
3. **Apply a statistical model to a large dataset to find outliers.**

The correct answer is; **Set up scheduled ingestion of data from the application databases to an analytical database**

Question **5**

Complete

Mark 1.00 out of 1.00

What is the correct Spark function to transform table in the left to the right?

Select one;

1. **df.extract()**
2. **df.groupby()**
3. **df.categorize()**

The correct answer is; **df.groupby()**

**Quiz Data Mrangling I** Jump to...

**Quiz Statistics for Data Science I €**



Copyright © 2019 IYKRA



Dashboard / My courses / Tsel\_DSA / Topic 21 ; Machine Learning IV

/ Quiz Machine Learning IV

**Started on** Friday, 24 January 2020, 6;36 PM

**State** Finished

**Completed on** Friday, 24 January 2020, 6;36 PM

**Time taken** 51 secs

**Marks** 5.00/5.00

**Grade 10.00** out of 10.00 (**100**%)

Question **1**

Complete

Mark 1.00 out of 1.00

Which is the method of validations?

Select one;

1. **k-Fold Cross Validation**
2. A,B, and C is correct
3. **Leave one out cross validation**
4. **Validation Set Approach**

The correct answer is; A,B, and C is correct

Question **2**

Complete

Mark 1.00 out of 1.00



How to make interactions for feature engineering, except?

Select one;

1. **Subset Selection**
2. **Addition**
3. **Multiplications**
4. **Substractions**

The correct answer is; **Subset Selection**

Question **S**

Complete

Mark 1.00 out of 1.00



What is the evaluation metric for classification?

Select one;

1. **R-squared**
2. **Mean squared error**
3. **MAPE**
4. **Precision**

The correct answer is; **Precision**

Question **4**



Complete

Mark 1.00 out of 1.00

What is the method that have the highest computational cost for feature selection?

Select one;

1. Backward Stepwise
2. Hybrid approach (Backward & Forward)
3. Best Subset Selection
4. Forward Stepwise

The correct answer is; **Best Subset Selection**

Question **5**

Complete

Mark 1.00 out of 1.00

How to measure %Event in WOE

Select one;

1. failed event in specific category Z all event in specific category
2. success event in specific category Z all event in specific category
3. success event in specific category Z success event in all category

The correct answer is; **success event in specific category Z all event in specific category**

# Quiz Data Solution and Architecture I

Jump to...

# Quiz Machine Learning V €

Copyright © 2019 IYKRA

Dashboard / My courses / Tsel\_DSA / Topic 22 ; Machine Learning V

/ Quiz Machine Learning V

**Started on** Friday, 24 January 2020, 6;30 PM



**State** Finished

**Completed on** Friday, 24 January 2020, 6;31 PM

**Time taken** 53 secs

**Marks** 5.00/5.00

**Grade 10.00** out of 10.00 (**100**%)

Question **1**

Complete

Mark 1.00 out of 1.00

These are supervised learning, except?

Select one;

1. Regression
2. Clustering
3. Classification

The correct answer is; **Clustering**

Question **2**

Complete

Mark 1.00 out of 1.00



What should we do if we have a lot of feature?

Select one;

1. **Feature Selection**
2. **Tuning Hyperparameter**
3. **Gain model complexity**
4. **Feature Engineering**

The correct answer is; **Feature Selection**

Question **S**

Complete

Mark 1.00 out of 1.00



Which is the evaluation metric for regression is?

Select one;

1. **F1-Score**
2. **R-squared**
3. **Accuracy**
4. **Logloss**

The correct answer is; **R-squared**

Question **4**



Complete

Mark 1.00 out of 1.00

These are application for classification, except?

Select one;

1. Face Recognition
2. Spam detection
3. Voice Recognition
4. Stock Prediction

The correct answer is; **Stock Prediction**

Question **5**

Complete

Mark 1.00 out of 1.00

If we have high bias, so?

Select one;

1. Model terlalu kompleks
2. Model terlalu sederhana
3. A dan C benar
4. Sensitif terhadap data baru

The correct answer is; **Model terlalu sederhana**

# Quiz Machine Learning IV

Jump to...

# Data Kaggle €

Copyright © 2019 IYKRA



Dashboard / My courses / Tsel\_DSA / Topic 23 ; Machine Learning VI

/ Quiz Machine Learning VI

**Started on** Friday, 24 January 2020, 6;25 PM

**State** Finished

**Completed on** Friday, 24 January 2020, 6;26 PM

**Time taken** 39 secs

**Marks** 5.00/5.00

**Grade 10.00** out of 10.00 (**100**%)

Question **1**

Complete

Mark 1.00 out of 1.00

Things should be considered in A/B Test? Select one;

a.

Hypotesis

b. All true c.

Measurement

d. Randomization

The correct answer is; All true

Question **2** Complete

Mark 1.00 out of 1.00



Which method can you use for regression ?

Select one;

1. All true
2. Decision Tree c.

XGBoost

d.

Random Forest

The correct answer is; All true

Question **S** Complete

Mark 1.00 out of 1.00



In linear regression, what transformation should you use when your target variable is higly right skewed ?

Select one;

a.

All true

b.

Quadratic transform(y2)

c. Log transform(log y) d.

Exponential transform(exp(y))

The correct answer is; Log transform(log y)

Question **4** Complete

Mark 1.00 out of 1.00



Metrics used in regression is ?

Select one;

a. Recall

b.

f1-score

c.

Precision

d. Mean squared error

The correct answer is; Mean squared error

Question **5** Complete

Mark 1.00 out of 1.00

What is the purpose of randomization ?

Select one;

a.

Remove selection bias

b. All true c.

Distributing co-variates evenly

The correct answer is; All true

**Data Kaggle** Jump to...

**Quiz Data solution & architecture II €**



Copyright © 2019 IYKRA



Dashboard / My courses / Tsel\_DSA / Topic 30 ; Scale up data science I

/ Quiz Scale up data science I

**Started on** Friday, 24 January 2020, 5;56 PM

**State** Finished

**Completed on** Friday, 24 January 2020, 5;57 PM

**Time taken** 43 secs

**Marks** 5.00/5.00

**Grade 10.00** out of 10.00 (**100**%)

Question **1**

Complete

Mark 1.00 out of 1.00

Jika model yang sudah kita train dan test tidak memenuhi standar akurasi yang ditetapkan, maka harus melakukan step ;

Select one;

1. Experimentation
2. Machine Learning Model and Tuning
3. Deploying Model
4. Gather Raw Data

The correct answer is; Machine Learning Model and Tuning

Question **2** Complete

Mark 1.00 out of 1.00



ekstensi .pkl adalah python object yang sudah melalui tahap

Select one;

1. Unit Testing
2. Error Check
3. Serialization
4. Debugging

The correct answer is; Serialization

Question **S** Complete

Mark 1.00 out of 1.00



Berikut adalah consideration dalam model deployment dan pembuatan analytics architecture, kecuali ;

Select one;

1. Reproducibility
2. Extensibility
3. Scalability
4. Visibility

The correct answer is; Visibility

Question **4** Complete



Mark 1.00 out of 1.00

Dibawah ini mana yang merupakan programming language yang disupport oleh Spark?

Select one;

1. Visual Basic
2. C++
3. Scala
4. C#

The correct answer is; Scala

Question **5** Complete

Mark 1.00 out of 1.00

Spark adalah

Select one;

1. High performance single computation system
2. High performance file storage
3. High performance hardware
4. High performance in-memory cluster computing

The correct answer is; High performance in-memory cluster computing

# Quiz Business Intelligence I

Jump to...

# Quiz Scale up data science II €

Copyright © 2019 IYKRA

Dashboard / My courses / Tsel\_DSA / Topic 31 ; Scale up data science II

/ Quiz Scale up data science II

**Started on** Friday, 24 January 2020, 5;39 PM



**State** Finished

**Completed on** Friday, 24 January 2020, 5;50 PM

**Time taken** 10 mins 54 secs

**Marks** 3.00/5.00

**Grade 6.00** out of 10.00 (**60**%)

Question **1**

Complete

Mark 1.00 out of 1.00

Berikut adalah components of scaling, kecuali

Select one;

1. Data handling
2. Preparing and fitting
3. Model training
4. Evaluation and deployment

The correct answer is; Preparing and fitting

Question **2** Complete

Mark 1.00 out of 1.00



Feature storage berguna untuk menyimpan

Select one;

1. variabel-variabel yang akan dievaluasi
2. variabel-variabel yang ada pada entry form
3. variabel-variabel yang akan digunakan untuk modelling
4. variabel-variabel yang ada pada aplikasi mobile

The correct answer is; variabel-variabel yang akan digunakan untuk modelling

Question **S** Complete

Mark 1.00 out of 1.00



Berikut adalah schema data processing

Select one;

1. Batch and out-time processing
2. Batch and real-time processing
3. Real-time and out-time processing
4. Real-time and at least once processing

The correct answer is; Batch and real-time processing

Question **4** Complete



Mark 0.00 out of 1.00

Serialisasi objek dalam bahasa pemrograman python akan berbentuk

Select one;

1. Series
2. Tuple
3. Pickle
4. PMML

The correct answer is; Pickle

Question **5** Complete

Mark 0.00 out of 1.00

Proses yang dilakukan terhadap feature sebelum masuk ke feature store adalah

Select one;

1. feature mining
2. feature blocking
3. feature engineering
4. feature piping

The correct answer is; feature engineering

# Quiz Scale up data science I

Jump to...

# Quiz Scale up data science III €

Copyright © 2019 IYKRA

Dashboard / My courses / Tsel\_DSA / Topic 32 ; Scale up data science III

/ Quiz Scale up data science III

**Started on** Friday, 24 January 2020, 5;34 PM



**State** Finished

**Completed on** Friday, 24 January 2020, 5;37 PM

**Time taken** 2 mins 42 secs

**Marks** 4.00/5.00

**Grade 8.00** out of 10.00 (**80**%)

Question **1**

Complete

Mark 1.00 out of 1.00

Berikut adalah contoh messaging service

Select one;

1. ActiveMQ
2. Zookeeper
3. Hadoop
4. Oracle

The correct answer is; ActiveMQ

Question **2** Complete

Mark 1.00 out of 1.00



Dua pihak yang berinteraksi di dalam kafka disebut

Select one;

1. Producer - Consumer
2. Consumer - Collaborator
3. Producer - Distributor
4. Distributor - Consumer

The correct answer is; Producer - Consumer

Question **S** Complete

Mark 1.00 out of 1.00



Producer bertugas untuk sebuah message

Select one;

1. Publish
2. Write-o"
3. Subscribe
4. Channel

The correct answer is; Publish

Question **4** Complete



Mark 0.00 out of 1.00

Zookeeper bertugas untuk

Select one;

1. memberihkan data dari kafka
2. membuat metadata dari kafka
3. menyimpan metadata dari kafka
4. menganalisis data dari kafka

The correct answer is; menyimpan metadata dari kafka

Question **5** Complete

Mark 1.00 out of 1.00

Manakah perusahaan yang menggunakan kafka pertama kali

Select one;

1. Linkedln
2. Spotify
3. Grab
4. Adidas

The correct answer is; Linkedln

# Quiz Scale up data science II

Jump to...

# Quiz Business Intelligence II €

Copyright © 2019 lYKRA



|  |  |  |
| --- | --- | --- |
| Dashboard / My courses / Tsel\_DSA / Topic 39 ; Soft Skills for data scientist III  / Quiz Soft Skills for data scientist III | | |
|  | **Started on** Thursday, 23 January 2020, 6;18 PM  **State** Finished  **Completed on** Thursday, 23 January 2020, 6;22 PM  **Time taken** 4 mins 19 secs  Question **1**  Complete  Marked out of 1.00  Berikut ini merupakan algoritma unsupervised learning yang benar adalah? Select one;   1. Clustering 2. Regression 3. Classification 4. Actor-critic algorithm |  |

Question **2** Complete

Marked out of 1.00



Berikut ini adalah contoh feature engineering yang termasuk dalam kategori basic adalah?

Select one;

1. Lag features
2. Count feature
3. Logarithm feature
4. Rolling windows feature

Question **S** Complete

Marked out of 1.00



Berikut ini adalah manfaat dari pengaplikasian cross validation yang benar adalah;

Select one;

1. Untuk mengetahui performance model terhadap test set
2. Jawaban A dan B benar
3. Untuk mengevaluasi performance dari model yang ada terhadap training set
4. Tidak ada jawaban yang benar

Question **4** Complete



Marked out of 1.00

Penentuan sebuah "Business Success Criteria" termasuk dalam scope metodologi CRISP-DM pada fase

Select one;

1. Evaluation
2. Data Understanding
3. Business Solution
4. Business Understanding

Question **5** Complete

Marked out of 1.00

Berikut adalah fakta yang benar mengenai project management Agile vs Waterfall

Select one;

1. Agile mempunyai peluang resiko lebih rendah, Waterfall lebih tinggi
2. Agile adalah fast-driven approach
3. Agile mempunyai peluang resiko project lebih tinggi, Waterfall lebih rendah
4. Waterfall adalah change-driven approach

**Quiz Data Science Delivery III**

Jump to...

Copyright © 2019 IYKRA



Dashboard / My courses / Tsel\_DSA / Topic 19 ; Soft Skills for DS I / Quiz Soft Skills for DS I

**Started on** Friday, 24 January 2020, 6;46 PM

**State** Finished

**Completed on** Friday, 24 January 2020, 6;46 PM

**Time taken** 37 secs

**Marks** 5.00/5.00

**Grade 10.00** out of 10.00 (**100**%)

Question **1**

Complete

Mark 1.00 out of 1.00

Below are example of business goals, expect...

Select one;

1. **Customer Satisfaction Improvement**
2. **Price Optimization**
3. **Fraud Prevention**
4. **Market Basket Analysis**

The correct answer is; **Market Basket Analysis**

Question **2**

Complete

Mark 1.00 out of 1.00



Which one is defined as Marketing Endgame?

Select one;

1. **CONVERSION**
2. **AWARENESS**
3. **ENGAGEMENT**
4. **TRANSACTION**

The correct answer is; **TRANSACTION**

Question **S**

Complete

Mark 1.00 out of 1.00



Which one IS NOT included in data solution proposal design?

Select one;

1. **Determine the approach**
2. **Define business background**
3. **None of the above**
4. **Data preprocessing**

The correct answer is; **Data preprocessing**

Question **4**



Complete

Mark 1.00 out of 1.00

Below are list of processes to achieve goals, except ...

Select one;

1. **Dynamic Pricing**
2. **Productivity Improvement**
3. **Customer Segmentation**
4. **Product Recommendation**

The correct answer is; **Productivity Improvement**

Question **5**

Complete

Mark 1.00 out of 1.00

Which one is less recommended to be included in your team when designing business solution?

Select one;

1. **Pricing Analyst**
2. **Data Project Manager**
3. **Business Analyst**
4. **Data Engineer**

The correct answer is; **Pricing Analyst**

# Quiz Machine Learning III

Jump to...

# Mid-Test SQL €

Copyright © 2019 IYKRA

Dashboard / My courses / Tsel\_DSA / Topic 13 ; Statistics for Data Science I

/ Quiz Statistics for Data Science I

**Started on** Friday, 24 January 2020, 7;15 PM

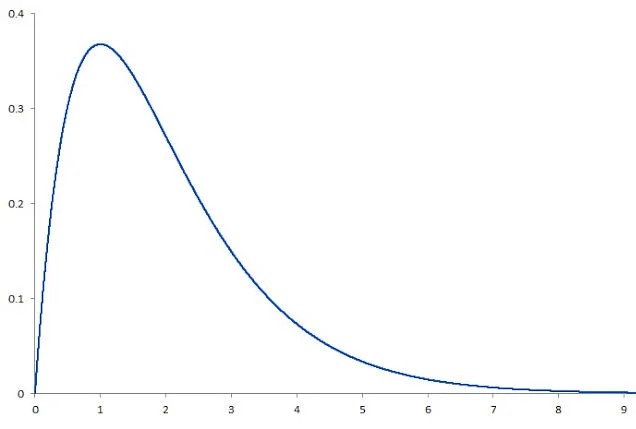
**State** Finished

**Completed on** Friday, 24 January 2020, 7;16 PM

**Time taken** 1 min 39 secs

**Marks** 7.00/7.00

**Grade 10.00** out of 10.00 (**100**%)



Question **1** Complete

Mark 1.00 out of 1.00

what term best to describe this chart?

Select one;

1. Left-skewed Distribution
2. Normal Distribution
3. Right-skewed Distribution

The correct answer is; Right-skewed Distribution

Question **2** Complete

Mark 1.00 out of 1.00



Number of children in one household is an example of ordinal variable.

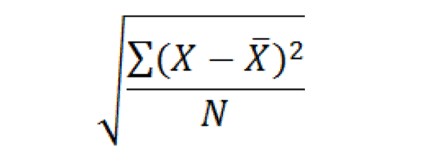
Select one;

1. False
2. True

The correct answer is; False

Question **S** Complete

Mark 1.00 out of 1.00



What is this formula for?

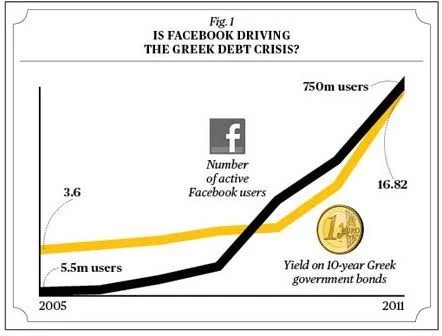
Select one;

1. Standard deviation
2. Median
3. Mean
4. Variance

The correct answer is; Standard deviation

Question **4** Complete

Mark 1.00 out of 1.00



Facebook growth caused the greek debt crisis.

Select one;

1. True
2. False

The correct answer is; False

Question **5** Complete

Mark 1.00 out of 1.00



The mode is the most appropriate measure of central tendency for which one of the following levels of measurement?

Select one;

1. Ordinal
2. Radio
3. Interval
4. Nominal

The correct answer is; Nominal

Question **6** Complete

Mark 1.00 out of 1.00



R squared in Regression measures ..

Select one;

1. the covariance between X and Y
2. the amount of variation in Y
3. the explained sum of squares as a proportion of the Total Sum Of Squares
4. the correlation between X and Y

The correct answer is; the explained sum of squares as a proportion of the Total Sum Of Squares

Question **7** Complete



Mark 1.00 out of 1.00

What is the probability of getting number 3 if we toss a fair dice once?

Select one;

a. 1/6

b. 2/6

c. 3/6

The correct answer is; 1/6

# Quiz Intro to Data Engineering and ETL Fundamentals using Python I

Jump to...

# Quiz Statistics for Data Science II €

Copyright © 2019 IYKRA

Dashboard / My courses / Tsel\_DSA / Topic 14 ; Statistics for Data Science II

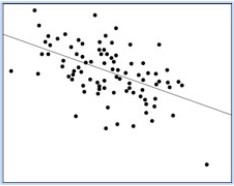
/ Quiz Statistics for Data Science II

|  |  |  |
| --- | --- | --- |
|  | **Started on** Friday, 24 January 2020, 7;08 PM  **State** Finished  **Completed on** Friday, 24 January 2020, 7;09 PM  **Time taken** 1 min 38 secs  **Marks** 5.00/5.00  **Grade 10.00** out of 10.00 (**100**%) |  |
|  |  |  |

Question **1**

Complete

Mark 1.00 out of 1.00



Which statement best describe coefficient of correlation of this scatter plot?

Select one;

a. **-1<R<0**

b. **0<R<1**

1. **R=0**
2. **R=1**
3. **R=-1**

The correct answer is; **-1<R<0**

Question **2**

Complete

Mark 1.00 out of 1.00



Which of the following coefficient of correlation indicates the strongest relationship

Select one;

a. **+0.58**

b. **+0.45**

c. **-0.67**

d. **-0.SS**

The correct answer is; **-0.67**

Question **S**

Complete

Mark 1.00 out of 1.00



Which is an example of Type I Error?

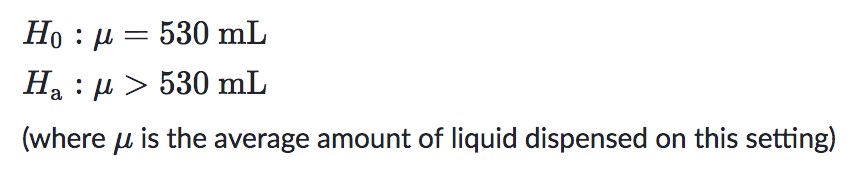
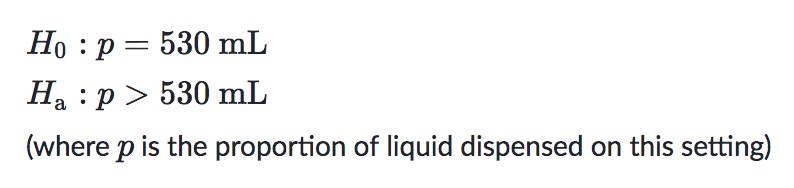
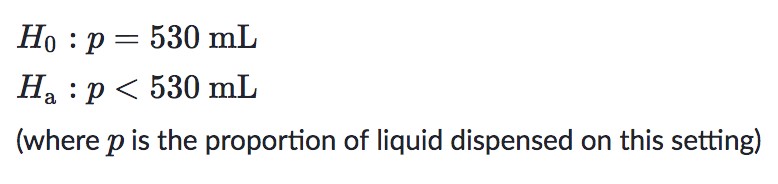
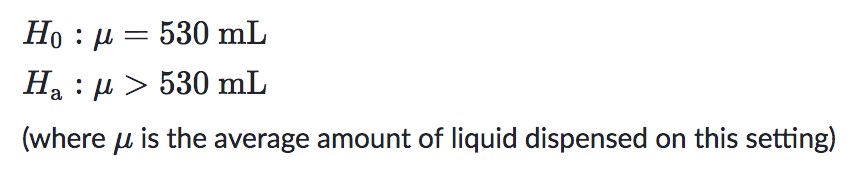
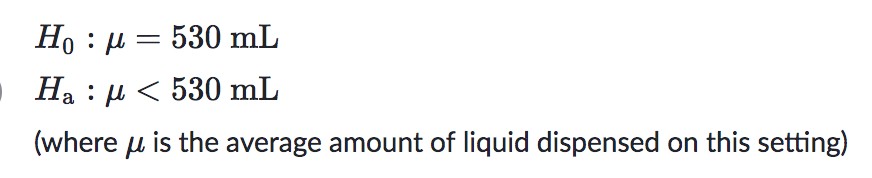
Select one;

1. **Falsely concluding there is an e"ect when actually there is no e"ect**
2. **Correctly concluding there is no e"ect**
3. **Falsely concluding there is no e"ect when actually there is an e"ect**
4. **Correctly concluding there is an e"ect**

The correct answer is; **Falsely concluding there is an e"ect when actually there is no e"ect**

Question **4** Complete

Mark 1.00 out of 1.00



The correct answer is;

A new automated drink machine is designed to dispense 530ml of liquid on the medium size setting. A restaurant owner suspects that his machine may be dispensing too much in medium drinks. She decide to take a sample of 303,030 medium drinks to see if the average amount is significantly greater than 530ml.

What are appropriate hypotheses for her significance test?

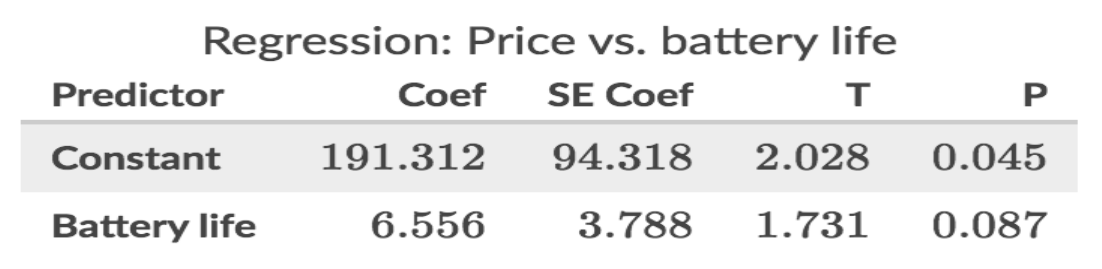
Select one;

a.

b.

c.

d.



Question **5**

Complete

Mark 1.00 out of 1.00

Using a significance level of alpha = 0.01, which of these is the most appropriate conclusion for this population of mobile phones?

Select one;

1. **This suggests a positive linear relationship between battery life and price because 0.045>0.01**
2. **We can't conclude a positive linear relationship between battery life and price because 0.045>0.01**
3. **We can't conclude a positive linear relationship between battery life and price because 0.087>0.01**
4. **This suggests a positive linear relationship between battery life and price because 0.087>0.01**

The correct answer is; **We can't conclude a positive linear relationship between battery life and price because 0.087>0.01**

# Quiz Statistics for Data Science I

Jump to...

# Quiz Statistics for Data Science III €

Copyright © 2019 IYKRA