Ce document a été mis en ligne par l'organisme FormaV®.

Toute reproduction, représentation ou diffusion, même partielle, sans autorisation préalable, est strictement interdite.

Pour en savoir plus sur nos formations disponibles, veuillez visiter :
formav.co/explorer

BACHELOR UNIVERSITAIRE DE TECHNOLOGIE

SCIENCE DES DONNÉES

ÉVALUATION DE COMPÉTENCE C5 : METTRE EN ŒUVRE UNE ANALYSE DESCRIPTIVE

SESSION 2026

Durée : 2 heures

SPÉCIALITÉ / UNITÉ	COEFF
SCIENCE DES DONNÉES	3,0 (COEFFICIENT INDICATIF)

Moyens de calculs et matériel autorisés :

L'usage de calculatrice avec mode examen actif est autorisé. L'usage de calculatrice sans mémoire, « type collège », est autorisé.

Aucun document ou matériel autorisé en dehors de ceux fournis avec le sujet.

Dès que le sujet vous est remis, assurez-vous qu'il est complet.

Ce sujet comporte 8 pages numérotées de 1/8 à 8/8.

Les documents réponses en fin de sujet sont à rendre avec la copie.

DOSSIER SUJET

MISE EN SITUATION PROFESSIONNELLE ET CONSIGNES GÉNÉRALES : ALTOMARCHÉ, SERVICE ÉTUDES ET PILOTAGE COMMERCIAL, LYON

Le service études d'AltoMarché souhaite décrire les commandes enregistrées sur une journée de vente afin d'éclairer une réunion de pilotage. Vous intervenez comme analyste de données. Le jeu fourni est un extrait anonymisé de 12 commandes. Il comporte le canal de vente, le montant du panier en euros et le délai de préparation en jours.

Le travail demandé porte sur une analyse descriptive : contrôler la nature des variables, calculer des indicateurs, comparer les canaux et formuler une restitution prudente. Les données de l'extrait ne constituent pas un échantillon aléatoire représentatif de l'ensemble de l'activité.

Les conventions de calcul, les données et les définitions nécessaires figurent dans le dossier technique. Les résultats numériques seront arrondis au centième lorsque cela est utile. La référence C5 et le coefficient indiqué correspondent à l'organisation pédagogique de l'évaluation.

PARTIE 1 : Cadrage des données et première description (5 points)

Identifier les variables, vérifier les unités et décrire les distributions

À partir des données du DT 1 et du dictionnaire du DT 2, établissez une première description de l'extrait. Justifiez vos choix à partir des définitions fournies.

Effectifs par canal de vente

Canal	Effectif	Fréquence en %
Web	.....	.....
Magasin	.....	.....
Total	.....	100

À compléter dans le document réponse DR 1.

1.1 (1,5 point) Pour chacune des trois variables du DT 1, indiquez sa nature statistique et son unité éventuelle. Précisez laquelle est une variable qualitative nominale et lesquelles sont quantitatives.

1.2 (2 points) Complétez le tableau du DR 1 avec les effectifs par canal et les fréquences en pourcentage. Utilisez la formule $f_i = \frac{n_i}{n} \times 100$, où $n = 12$.

1.3 (1,5 point) Décrivez la distribution des délais de préparation en indiquant l'effectif de chaque délai observé, de 1 à 5 jours. Quel est le délai modal ?

PARTIE 2 : Indicateurs descriptifs des montants de panier (6 points)

Calculer et interpréter des indicateurs de position et de dispersion

La variable étudiée est le montant du panier, exprimé en euros. Utilisez les conventions de calcul du DT 2, notamment la variance descriptive avec division par n et la définition des quartiles par rang supérieur.

Valeurs ordonnées des montants de panier

Rang	1	2	3	4	5	6	7	8	9	10	11	12
Montant (€)	38	42	44	46	49	51	55	58	62	70	75	90

Les montants sont exprimés en euros. La liste est fournie pour faciliter les calculs.

2.1 (2 points) Calculez la moyenne et la médiane des montants de panier. Présentez une formule, les valeurs utilisées et le résultat en euros, arrondi au centième si nécessaire.

2.2 (1,5 point) Déterminez le premier quartile Q_1, le troisième quartile Q_3 et l'écart interquartile. Indiquez les rangs retenus conformément à la convention du DT 2.

2.3 (2,5 points) Calculez la variance descriptive et l'écart-type des montants. Vous pouvez utiliser $\sum x_i = 680$ et $\sum x_i^2 = 41\,120$, valeurs obtenues à partir du DT 1. Donnez les unités pertinentes et interprétez brièvement l'écart-type dans le contexte.

PARTIE 3 : Comparaison descriptive des canaux de vente (5 points)

Construire une comparaison lisible et interpréter sans généraliser abusivement

Le responsable souhaite savoir si les montants observés diffèrent entre les commandes Web et les commandes en magasin. Calculez les indicateurs demandés à partir des lignes du DT 1, puis formulez une comparaison descriptive limitée à cet extrait.

Tableau de synthèse à compléter

Canal	Effectif	Somme des paniers (€)	Moyenne des paniers (€)	Délai moyen (jours)
Web	.....	.....	.....	.....
Magasin	.....	.....	.....	.....

Les calculs portent sur les 12 commandes de l'extrait.

3.1 (2,5 points) Complétez le tableau du DR 1 en calculant, pour chaque canal, l'effectif, la somme et la moyenne des montants de panier, ainsi que le délai moyen de préparation. Donnez les moyennes au centième si nécessaire.

3.2 (1,5 point) Rédigez une comparaison de deux ou trois phrases entre les canaux à partir des résultats calculés. Distinguez explicitement le constat portant sur l'extrait d'une conclusion qui concernerait l'ensemble des clients.

3.3 (1 point) Proposez une représentation graphique adaptée pour comparer les montants selon le canal. Précisez les variables portées par les axes ou les modalités représentées et justifiez brièvement votre choix.

PARTIE 4 : Valeurs atypiques, qualité des données et restitution (4 points)

Contrôler une observation et communiquer des résultats avec prudence

Avant la réunion, vous devez vérifier si une valeur de panier mérite un contrôle et proposer une restitution adaptée au responsable. Les seuils de détection demandés sont ceux indiqués dans le DT 2.

4.1 (1,5 point) À l'aide de Q_1, Q_3 et de l'écart interquartile calculés en partie 2, calculez les bornes $Q_1 - 1{,}5 \times EIQ$ et $Q_3 + 1{,}5 \times EIQ$. Indiquez si une valeur de l'extrait est signalée par cette règle.

4.2 (1 point) Expliquez pourquoi une valeur signalée comme atypique ne doit pas être supprimée automatiquement. Donnez deux vérifications concrètes à effectuer avant toute décision de correction ou d'exclusion.

4.3 (1,5 point) Rédigez une synthèse de trois phrases au maximum destinée au responsable. Elle doit présenter un résultat descriptif chiffré, une limite liée à la taille ou à la constitution de l'extrait et une précaution de confidentialité ou de qualité des données.

DOSSIER TECHNIQUE ET RESSOURCES

Documents ressources nécessaires à la résolution du sujet

DT 1 : EXTRAIT ANONYMISÉ DES COMMANDES

Chaque ligne correspond à une commande. Les identifiants sont fictifs et ne permettent pas d'identifier un client. Les montants sont en euros et les délais sont en jours entiers. L'extrait comprend 12 commandes enregistrées sur une journée de vente.

Données observées

Commande	Canal	Montant du panier (€)	Délai de préparation (jours)
O01	Web	42	2
O02	Web	55	3
O03	Magasin	38	1
O04	Web	70	4
O05	Magasin	46	2
O06	Web	62	3
O07	Magasin	51	2
O08	Web	90	5
O09	Magasin	44	1
O10	Web	58	3
O11	Magasin	75	4
O12	Web	49	2

Source : extraction de démonstration du service études AltoMarché, identifiants anonymisés.

DT 2 : DICTIONNAIRE DES VARIABLES ET CONVENTIONS DE CALCUL

Le canal est une variable qualitative nominale à deux modalités : Web et Magasin. Le montant du panier est une variable quantitative exprimée en euros. Le délai de préparation est une variable quantitative discrète exprimée en jours.

Pour la fréquence, utiliser $f_i = \frac{n_i}{n}$. Pour la moyenne, utiliser $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Pour la variance descriptive, utiliser $V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$, puis $s = \sqrt{V}$. Ne pas utiliser la variance corrigée avec dénominateur $n-1$.

La médiane d'une série ordonnée de 12 valeurs est la moyenne des valeurs de rangs 6 et 7. Pour un quartile de niveau p, le rang retenu est $\lceil pn \rceil$, où $\lceil \cdot \rceil$ désigne l'entier supérieur ou égal. Ainsi, Q_1 est de rang $\lceil 0{,}25n \rceil$ et Q_3 de rang $\lceil 0{,}75n \rceil$. L'écart interquartile est $EIQ = Q_3 - Q_1$.

Pour le contrôle demandé en partie 4, une observation est signalée si elle est inférieure à $Q_1 - 1{,}5 \times EIQ$ ou supérieure à $Q_3 + 1{,}5 \times EIQ$. Ce signalement est un indicateur de contrôle, pas une preuve d'erreur.

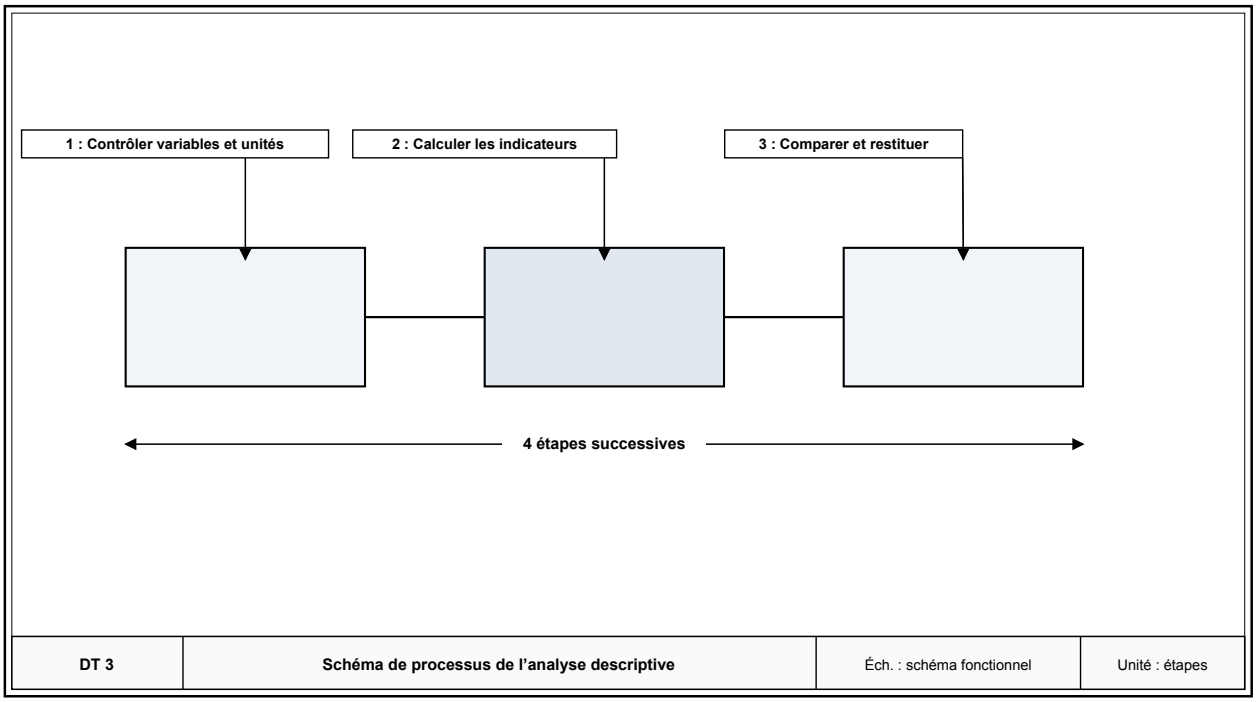
Repères de calcul fournis

Grandeur	Valeur
Effectif total	12
Somme des montants	680 €
Somme des carrés des montants	41 120 €²
Minimum et maximum des montants	38 € et 90 €

Valeurs calculées sur les 12 montants du DT 1.

DT 3 : CIRCUIT DE PRODUCTION D'UNE SYNTHÈSE DESCRIPTIVE

Le schéma présente le circuit de travail demandé dans cette étude de cas : contrôler les variables et les unités, calculer les indicateurs, comparer les groupes, puis restituer les résultats avec leurs limites. Il s'agit d'un schéma de processus d'analyse, et non d'un plan d'ouvrage ou d'une prescription réglementaire.



NE RIEN ÉCRIRE DANS CETTE PARTIE

DR 1 : Document réponse : Questions 1.2, 3.1 et 3.2 — **À RENDRE AVEC LA COPIE**

Complétez les tableaux directement. Indiquez les unités et présentez les calculs utiles dans votre copie. Pour les fréquences, arrondissez au dixième de pourcentage si nécessaire.

Répartition des commandes par canal

Canal	Effectif	Fréquence en %
Web	.....	.....
Magasin	.....	.....
Total	.....	100

À compléter pour la question 1.2.

Indicateurs descriptifs par canal

Canal	Effectif	Somme des paniers (€)	Moyenne des paniers (€)	Délai moyen (jours)
Web	.....	.....	.....	.....
Magasin	.....	.....	.....	.....

À compléter pour la question 3.1. Arrondir les moyennes au centième si nécessaire.

Cadre réservé aux justifications, calculs complémentaires ou schémas du candidat :

.....

.....

.....

.....

.....

.....

.....

NE RIEN ÉCRIRE DANS CETTE PARTIE

DR 2 : Document réponse : Synthèse destinée au responsable — **À RENDRE AVEC LA COPIE**

Rédigez votre synthèse dans la zone de réponse de votre copie en respectant les trois exigences de la question 4.3 : résultat chiffré, limite de portée et précaution de confidentialité ou de qualité des données.

Plan de préparation de la restitution

Élément attendu	Repère à noter
Résultat descriptif chiffré	.....
Limite de portée	.....
Précaution de qualité ou de confidentialité	.....

Repères facultatifs à renseigner avant de rédiger la synthèse.

Cadre réservé aux justifications, calculs complémentaires ou schémas du candidat :

.....

.....

.....

.....

.....

.....

.....

Copyright © 2027 FormaV®. Tous droits réservés.

FormaV® est une marque déposée auprès de l'INPI sous le numéro 5227897.

Ce document a été élaboré par FormaV® avec le plus grand soin afin d'accompagner chaque apprenant vers la réussite de ses examens. Son contenu (textes, graphiques, méthodologies, tableaux, exercices, concepts, mises en forme) constitue une œuvre protégée par le droit d'auteur.

Toute copie, partage, reproduction, diffusion ou mise à disposition, même partielle, gratuite ou payante, est strictement interdite sans accord préalable et écrit de FormaV®, conformément aux articles L.111-1 et suivants du Code de la propriété intellectuelle. Dans une logique anti-plagiat, FormaV® se réserve le droit de vérifier toute utilisation illicite, y compris sur les plateformes en ligne ou sites tiers.

En utilisant ce document, vous vous engagez à respecter ces règles et à préserver l'intégrité du travail fourni. La consultation de ce document est strictement personnelle.

Merci de respecter le travail accompli afin de permettre la création continue de ressources pédagogiques fiables et accessibles.