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

BUT SD

BUT Science des Données : compétence C1

Session 2026 • Code sujet : **26-BUT-SD-EP1** • Durée : **2 heures, durée retenue pour cette proposition de sujet** • Coefficient : **3,0 indicatif**

Sommaire

Cliquez sur une partie pour y accéder directement

1. **Analyse du besoin et modélisation des données** (3 questions) — **Page 3**
2. **Requêtes SQL et cycle de vie des données** (3 questions) — **Page 5**
3. **Automatisation du traitement et production d'indicateurs** (3 questions) — **Page 7**
4. **Prévention des risques et documentation** (3 questions) — **Page 9**
5. **Ce qu'il faut retenir & Questions fréquentes (FAQ)** (Synthèse & Q&A Examen) — **Page 11**

Sujet officiel (PDF) : https://butsd.fr/assets/annales/but-sd/C1 - Traiter des données à des fins décisionnelles/2026/BUT Science des Données - C1 - Traiter des données à des fins décisionnelles - Session 2026 - 1-Sujet.pdf

1. Analyse du besoin et modélisation des données

La Direction des bâtiments de Valmont-sur-Rive veut fiabiliser le suivi électrique de trois bâtiments municipaux à partir des relevés de mars 2026. Il faut organiser les données en relations et prévoir des contrôles avant leur chargement.

Modèle relationnel

Question 1.1. Proposez un modèle relationnel minimal adapté au besoin. Indiquez les relations, leurs clés primaires et la clé étrangère, les types de données adaptés aux champs principaux ainsi que la cardinalité entre bâtiment et relevé.

Le modèle distingue les bâtiments, données de référence, des relevés de consommation. **Batiment**(id_batiment `VARCHAR`, nom `VARCHAR`, surface_m2 `DECIMAL`) a pour clé primaire `id_batiment`. **Releve**(id_releve `VARCHAR`, id_batiment `VARCHAR`, date_releve `DATE`, conso_kwh `DECIMAL`, source `VARCHAR`) a pour clé primaire `id_releve` et pour clé étrangère `id_batiment`, non nulle, référençant `Batiment.id_batiment`.

Un bâtiment peut être associé à zéro, un ou plusieurs relevés. Chaque relevé concerne exactement un bâtiment.

> ✓ RÉPONSE À RETENIR
>
> **Batiment**(id_batiment PK, nom, surface_m2) ; **Releve**(id_releve PK, id_batiment FK, date_releve, conso_kwh, source). Cardinalité : un bâtiment pour 0..N relevés.

Séparation et validation des données

Question 1.2. Expliquez en quoi la séparation des bâtiments et des relevés évite une anomalie de mise à jour. Donnez une règle de validation pertinente pour la date et une règle pertinente pour la consommation.

La séparation évite de répéter le nom et la surface d'un bâtiment dans chaque relevé. Si ces informations changent, une seule ligne de **Batiment** est mise à jour : on évite ainsi les oublis et les valeurs contradictoires dans les relevés associés.

Pour `date_releve`, vérifier que la valeur n'est pas nulle et correspond à une date valide. Pour `conso_kwh`, vérifier qu'elle est numérique, non nulle et supérieure ou égale à zéro, conformément aux règles de qualité du sujet.

> ✓ RÉPONSE À RETENIR
>
> La séparation évite les mises à jour répétées et incohérentes. Valider une date réelle non nulle et une consommation numérique non nulle, supérieure ou égale à zéro.

Contrôles avant chargement

Question 1.3. Complétez le DR 1 avec trois contrôles à effectuer avant le chargement en base. Pour chaque contrôle, indiquez la donnée concernée et l'action à mener si le contrôle échoue.

Les trois contrôles ci-dessous complètent le DR 1. Ils vérifient la validité de la date, l'intégrité des identifiants et la cohérence de la consommation avant le chargement. Si une ligne échoue, elle est mise à l'écart du chargement jusqu'à vérification ou correction. Toute correction doit être tracée avec l'identifiant du relevé, l'ancienne et la

nouvelle valeur, la date et l'auteur.

DR 1 complété : contrôles avant chargement

Alerte	Consommation strictement supérieure à 320 kWh	
P	indicateurs	5 mars 2026 inclus
Date valide	**date_releve non nulle, date réelle**	**Rejeter la ligne ; faire corriger la source**
Identifiants	**id_releve unique et non nul ; id_batiment existant**	**Écarter la ligne ; vérifier les références**
Consommation	**conso_kwh numérique, non nulle et ≥ 0**	**Mettre en attente ; faire vérifier la valeur**

Règles définies par le cahier des charges de la Direction des bâtiments.

✓ RÉPONSE À RETENIR

Vérifier date valide, identifiant unique et référence à un bâtiment existant, puis consommation numérique et supérieure ou égale à zéro. Mettre en attente les lignes non conformes.

2. Requêtes SQL et cycle de vie des données

Le dossier présente les relations Batiment et Releve et les relevés de mars 2026, avec un doublon à supprimer, une consommation à corriger et un enregistrement à insérer. Les requêtes portent sur la période du 1er au 15 mars inclus et sur le seuil d'alerte strictement supérieur à 320 kWh.

Nettoyage et insertion

Question 2.1. Écrivez, dans l'ordre demandé, les instructions SQL pour supprimer le doublon identifié au DT 2, corriger la consommation du relevé R006 à 235 kWh, puis insérer le relevé R010 indiqué au DT 2. Utilisez des conditions suffisamment précises pour ne modifier que les enregistrements visés.

Effectuer les trois opérations dans l'ordre indiqué. La suppression cible R000, doublon exact de R001. La correction vérifie aussi les valeurs initiales de R006. R010 est ensuite ajouté avec les données exactes du DT 2; sa date le place hors de la période des indicateurs.

```
DELETE FROM Releve
WHERE id_releve = 'R000'
  AND id_batiment = 'S01'
  AND date_releve = DATE '2026-03-01'
  AND conso_kwh = 420
  AND source = 'compteur';

UPDATE Releve
SET conso_kwh = 235
WHERE id_releve = 'R006'
  AND id_batiment = 'S03'
  AND date_releve = DATE '2026-03-08'
  AND conso_kwh = 230
  AND source = 'compteur';

INSERT INTO Releve (id_releve, id_batiment, date_releve, conso_kwh, source)
VALUES ('R010', 'S01', DATE '2026-03-22', 400, 'compteur');
```

> ✓ **RÉPONSE À RETENIR**
>
> R000 supprimé; R006 corrigé à 235 kWh; R010 inséré.

Consommation moyenne

Question 2.2. Écrivez une requête SQL qui retourne, pour chaque bâtiment, son nom et la consommation moyenne des relevés valides du 1er au 15 mars 2026 inclus. Arrondissez la moyenne à deux décimales, triez le résultat par identifiant de bâtiment croissant et excluez le doublon après nettoyage.

La borne supérieure exclusive au 16 mars inclut bien les relevés jusqu'au 15 mars. La moyenne est calculée sur les lignes après nettoyage: R000 est supprimé et R006 est corrigé à 235 kWh. Le regroupement par bâtiment et le tri par identifiant produisent les résultats dans l'ordre demandé.

```
SELECT b.nom,
       ROUND(AVG(r.conso_kwh), 2) AS consommation_moyenne_kwh
FROM Batiment AS b
JOIN Releve AS r ON r.id_batiment = b.id_batiment
WHERE r.date_releve >= DATE '2026-03-01'
  AND r.date_releve < DATE '2026-03-16'
GROUP BY b.id_batiment, b.nom
ORDER BY b.id_batiment ASC;
```

✓ RÉPONSE À RETENIR

Centre administratif : 405,00 kWh; Gymnase municipal : 321,67 kWh; Médiathèque : 241,67 kWh.

Comptage des alertes

Question 2.3. Écrivez une requête SQL qui compte les relevés en alerte sur la même période, après nettoyage. Indiquez le résultat attendu à partir des données du DT 2.

Le filtre reprend la période du 1er au 15 mars inclus et applique le seuil strict: seuls les relevés supérieurs à 320 kWh sont comptés. Après suppression de R000, les relevés en alerte sont R001, R002, R004, R007 et R008. R010 est exclu car daté du 22 mars.

```
SELECT COUNT(*) AS nombre_alertes
FROM Releve
WHERE date_releve >= DATE '2026-03-01'
  AND date_releve < DATE '2026-03-16'
  AND conso_kwh > 320;
```

✓ RÉPONSE À RETENIR

5 relevés en alerte.

3. Automatisation du traitement et production d’indicateurs

Le traitement automatisé utilise les relevés corrigés et dédoublonnés du 1er au 15 mars 2026 inclus ; R010, daté du 22 mars, est exclu. Il faut programmer une fonction de calcul, produire des indices surfaciques et des évolutions, puis proposer des tests unitaires.

Fonction de calcul

Question 3.1. Rédigez en Python ou en pseudo-code structuré une fonction qui reçoit les consommations d’un bâtiment et sa surface, puis renvoie la consommation totale rapportée à la surface. La fonction doit rejeter une liste vide et une surface inférieure ou égale à zéro. Commentez les contrôles réalisés.

La fonction refuse une liste vide, une surface non numérique ou non positive, et des consommations non numériques ou négatives. Ces contrôles respectent les règles de qualité du sujet. Elle renvoie ensuite la somme des consommations divisée par la surface.

```
def indice_surfacique(consommations, surface):
    if not consommations:
        raise ValueError("Liste vide")
    if not isinstance(surface, (int, float)) or surface <= 0:
        raise ValueError("Surface non positive ou invalide")
    if any(not isinstance(c, (int, float)) or c < 0
           for c in consommations):
        raise ValueError("Consommation négative ou invalide")
    return sum(consommations) / surface
```

> ✓ RÉPONSE À RETENIR
>
> Indice surfacique = somme des consommations / surface, avec rejet des entrées invalides.

Calcul des indicateurs

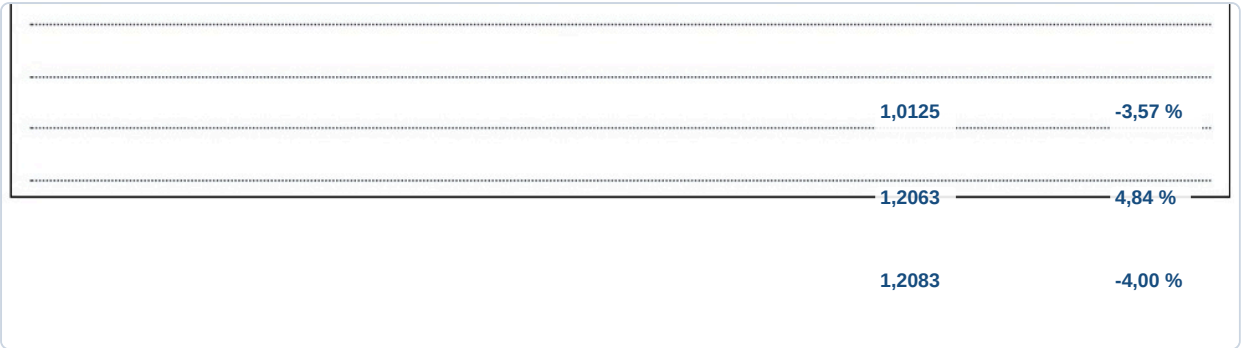
Question 3.2. Calculez les trois indices en kWh/m² et les trois évolutions en pourcentage, en utilisant pour chaque bâtiment le premier et le dernier relevé valide de la période. Arrondissez les indices à quatre décimales et les évolutions à deux décimales, puis complétez le DR 2.

Les indices sont calculés avec les consommations corrigées et les surfaces. Les évolutions utilisent le premier et le dernier relevé valide du 1er au 15 mars ; R010 est exclu. Après correction de R006 à 235 kWh, le total de S03 est 725 kWh, et non 715 kWh comme indiqué dans le tableau prérempli du sujet.

Bâtiment	Indice (kWh/m²)	Évolution (%)
S01	1,0125	-3,57
S02	1,2063	4,84
S03	1,2083	-4,00

Pour S03, l’évolution reste calculée entre R005 (250 kWh) et R009 (240 kWh).

DR 2 complété



✓ RÉPONSE À RETENIR

S01 : 1,0125 kWh/m², -3,57 % ; S02 : 1,2063 kWh/m², 4,84 % ; S03 : 1,2083 kWh/m², -4,00 %.

Tests unitaires

Question 3.3. Proposez deux tests unitaires vérifiables pour la fonction de la question 3.1, dont un test de cas limite. Indiquez pour chacun les données d'entrée et le résultat attendu.

Le premier test vérifie un calcul avec deux consommations. Le second est un cas limite valide : une consommation nulle, autorisée par les règles du sujet.

1. Entrées : consommations `[100, 200]`, surface `100`. Résultat attendu : `3.0`.
2. Cas limite : consommations `[0]`, surface `5`. Résultat attendu : `0.0`.

✓ RÉPONSE À RETENIR

([100, 200], 100) donne 3,0 ; ([0], 5) donne 0,0 kWh/m².

4. Prévention des risques et documentation

Les indicateurs seront transmis à la Direction des bâtiments dans un tableau de bord mensuel, tandis que les fichiers sources et les résultats sont conservés dans l'espace professionnel de la collectivité. Cette partie évalue la prévention des risques, la documentation destinée au métier et la reproductibilité du traitement.

Prévention des risques

Question 4.1. Pour chacune des trois situations du tableau, proposez une mesure de prévention ou de réduction du risque. Votre réponse doit inclure une mesure portant sur les droits d'accès, une mesure portant sur la sauvegarde ou la traçabilité et une mesure portant sur la validation des données.

Situation	Mesure de prévention
Fichier source remplacé	Conserver une copie datée, non modifiable, et tracer chaque version.
Modification par un utilisateur non habilité	Réserver les droits d'écriture aux personnes autorisées et journaliser les modifications.
Tableau de bord incomplet	Vérifier la présence de la période et des unités avant d'autoriser la diffusion.

Les versions sauvegardées assurent la traçabilité, les droits limitent les altérations et la validation prévient les interprétations erronées.

> ✓ RÉPONSE À RETENIR
>
> Versions datées sauvegardées, accès d'écriture restreints et tableau de bord vérifié avant diffusion.

Documentation métier

Question 4.2. Rédigez une courte documentation destinée à un responsable métier. Elle doit préciser la période étudiée, la définition de l'alerte, l'unité de l'indice surfacique et le sens d'une évolution négative.

Le tableau de bord porte sur les relevés valides du 1er au 15 mars 2026 inclus. Un relevé est en alerte si sa consommation dépasse strictement 320 kWh. L'indice surfacique est la consommation totale divisée par la surface du bâtiment, exprimée en kWh/m². L'évolution compare le dernier relevé au premier de la période : une valeur négative signifie que la consommation du dernier relevé est inférieure à celle du premier.

> ✓ RÉPONSE À RETENIR
>
> Période : 1er-15 mars 2026 inclus ; alerte : consommation > 320 kWh ; indice : kWh/m² ; évolution négative : baisse entre premier et dernier relevé.

Conservation et reproductibilité

Question 4.3. Expliquez pourquoi le traitement doit conserver séparément les données sources, les données corrigées et les résultats calculés. Indiquez un moyen concret de rendre le calcul reproductible par une autre personne de l'équipe.

Les sources sont conservées sans modification pour garder une référence fidèle aux fichiers reçus. Les corrections sont séparées et tracées avec l'identifiant du relevé, l'ancienne et la nouvelle valeur, la date et l'auteur. Les résultats calculés restent distincts des données d'entrée. Pour permettre à une autre personne de reproduire le calcul, conserver un script versionné avec les règles appliquées et les données corrigées utilisées.

> **✓ RÉPONSE À RETENIR**
>
> Conserver séparément sources, corrections tracées et résultats ; versionner le script et ses règles pour reproduire le calcul.

Ce qu'il faut retenir de ce sujet (Session 2026)

Bilan de l'épreuve : Cette proposition d'évaluation de la compétence C1 du BUT Science des Données est celle de la session 2026, code 26-BUT-SD-EP1, pour une durée de 2 heures et un coefficient indicatif de 3,0. À partir des consommations électriques de trois bâtiments municipaux en mars, elle évalue la modélisation relationnelle, les contrôles qualité, SQL, le calcul d'indicateurs, la programmation, les tests, la prévention des risques et la documentation.

Notions-clés, formules & règles à maîtriser

- Chaque relevé concerne un seul bâtiment et sa clé étrangère doit référencer un identifiant existant dans la relation Batiment.
- Une alerte correspond à une consommation strictement supérieure à 320 kWh ; 320 kWh exactement ne déclenche pas d'alerte.
- L'indice surfacique est la consommation totale en kWh divisée par la surface en m² ; son unité est le kWh/m².
- L'évolution relative est T = 100 × (Cdernier - Cpremier) / Cpremier ; une valeur négative signifie que la dernière consommation est inférieure à la première.

Points de vigilance & attentes du jury

- Distinguer clés primaires et étrangère, préciser les types adaptés et respecter la cardinalité bâtiment-relevé.
- Supprimer le doublon, corriger R006 puis insérer R010 dans l'ordre demandé, avec des conditions SQL ciblées.
- Retenir les dates du 1er au 15 mars 2026 inclus ; R010, daté du 22 mars, est hors période.
- Respecter les arrondis demandés et vérifier les contrôles d'entrée ainsi qu'un cas limite dans les tests.

Questions fréquentes sur cette épreuve (Q&A Examen)

Q : Comment calculer l'évolution entre le premier et le dernier relevé ?

Soustrayez la consommation du premier relevé de celle du dernier, divisez la différence par la première consommation, puis multipliez par 100. Une valeur négative signifie que la dernière consommation est plus faible.

Q : Quelle période faut-il retenir pour les requêtes et les indicateurs ?

La période comprend les relevés du 1er au 15 mars 2026 inclus. En SQL, utilisez une date supérieure ou égale au 1er mars et strictement antérieure au 16 mars. Le relevé R010 est exclu.

Q : Combien de points et de questions comporte le sujet ?

Le sujet comporte quatre parties de 5 points, soit 20 points au total, et douze questions numérotées de 1.1 à 4.3. L'énoncé n'impose aucune répartition du temps entre les parties.

Q : Quels arrondis et quel seuil d'alerte faut-il respecter ?

Arrondissez les indices surfaciques à quatre décimales et les évolutions à deux décimales. Le seuil est strict : seule une consommation supérieure à 320 kWh déclenche une alerte, pas une valeur égale à 320.

Sujet officiel associé (PDF) : https://butsd.fr/assets/annales/but-sd/C1 - Traiter des données à des fins décisionnelles/2026/BUT Science des Données - C1 - Traiter des données à des fins décisionnelles - Session 2026 - 1-Sujet.pdf
Ressources BUT Science des Données : Toutes les annales & corrigés • Fiches de révision complètes • Simulateur de notes d'examen

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.