

Examen blanc: Analyse de données

Partie I : Analyse de Données dans le Secteur du E-commerce

Contexte :

Une entreprise de commerce électronique, "TechStyle", vend des produits de mode en ligne. Au cours des dernières années, l'entreprise a recueilli de nombreuses données concernant les commandes, les clients et leurs produits. L'entreprise a remarqué qu'il y a des fluctuations de ventes selon les mois. Elle souhaite donc analyser ces données afin de :

1. Mieux comprendre le comportement d'achat de ses clients.
2. Identifier les facteurs qui influencent le niveau des ventes et les produits les plus populaires.
3. Optimiser ses stratégies de marketing et de gestion des stocks.

Tableau de Distribution Empirique : Montant des Commandes (en euros)

Tranche de Montant de Commandes (€)	Fréquence Absolue (Nombre de Commandes)	Fréquence Relative (%)
0 - 50	3	15%
51 - 100	2	10%
101 - 150	4	20%
151 - 200	4	20%
201 - 250	4	20%
251 - 300	2	10%
301 +	1	5%
Total	20	100%

A. Analyse univariée

Questions :

1. **Mesure de tendance centrale** : À partir du tableau de distribution empirique ci-dessus, vous allez calculer :
 - Estimez la moyenne approximative du montant de commande.
 - Déterminez la médiane. Comment interprétez-vous cette valeur?
 - Quel serait le mode de cette distribution? Que vous indique t-il sur la fréquence des montants dans cette table de données?

2. **Mesures de dispersion** : En utilisant les données du tableau,

- Estimez la dispersion des valeurs des montants des commandes en calculant l'étendue. Expliquez cette valeur.
- Expliquez ce que sont les quartiles, et estimez la valeur du premier quartile (Q1) et le troisième quartile (Q3) en utilisant votre distribution empirique (donnez votre démarche), et expliquez leur utilité.
- Estimez, en utilisant les valeurs précédemment calculées, l'écart interquartile (IQR), et expliquez en quoi cette mesure permet de compléter l'information sur la dispersion de la distribution.

3. **Variance et Écart-Type** :

4. **Visualisation et Interprétation** :

- Si vous deviez afficher ces informations sur la distribution du montant des commandes avec un histogramme, quelle sera la forme (distribution symétrique ou asymétrique) et la disposition attendue de ce graphique. Quels types de conclusions pourrait-on tirer, en comparant vos résultats calculés aux points précédents avec votre graphe.
- Dans le cas d'une représentation sous forme de box plot, identifiez et interprétez sur un box plot (que vous décrirez) la médiane, les quartiles, les moustaches, les valeurs aberrantes (si elles existent). Comment pouvez vous relier les résultats observés sur ce boxplot avec la valeur des indicateurs calculés précédemment?

B. Analyse Bivariée

Tableau : Prix des Commandes par Niveau d'Activité des Clients

Niveau d'Activité	Liste des Prix par Commandes (€)
Faible	150, 200, 100, 250, 300, 250
Modéré	120, 170, 230, 110, 240, 290
Intense	150, 230, 145, 220, 240, 255

Questions :

Justifiez le choix d'un test statistique approprié pour la comparaison de plusieurs moyennes (par exemple: l'ANOVA, et mettre en avant ses conditions d'utilisation). Vous préciserez les hypothèses du test et comment la p-value est utilisée pour conclure.

* Mettez en place votre test et interprétez vos résultats. Donnez la p-value et indiquez la conclusion en se basant sur un seuil de confiance de 5%

Tableau : Lien entre catégorie et motif de retour :

Motif_Retour	Vêtements	Accessoires	Chaussures	Total
Taille	25	5	2	32
Couleur	10	10	3	23

Défectueux	5	7	10	22
Autre	10	3	10	23
Total	50	25	25	100

* Réaliser un test du Chi-2. Vous indiquerez la nature des tests, les hypothèses de bases.

Interprétation : Que peut-on conclure sur la dépendance ou l'indépendance entre la catégorie des produits et le type de retour ?

C. Visualisation et corrélation :

Objectif : Étudiez, si la variable d'évaluation (Score Moyen) influence de manière significative, le prix d'un produit et si c'est le cas, préciser l'amplitude et le sens de cette relation.

Voici un tableau simple avec deux colonnes, suivi de questions pour guider l'analyse de corrélation, causalité et régression linéaire :

Produit	Score Moyen	Prix (€)
Produit A	4.5	50
Produit B	3.8	30
Produit C	4.0	35
Produit D	2.5	15
Produit E	4.7	60
Produit F	3.2	20
Produit G	4.9	80

Questions :

1. Corrélation

- Utilisez un graphique adapté, en justifiant votre choix (nuage de points).
- Calculez le coefficient de corrélation entre le **Score Moyen** et le **Prix (€)**.
- Interprétez la valeur du coefficient de corrélation en la reliant avec ce que vous observez dans votre graphique (force et direction).
- Interprétez la force et le sens de cette relation : est-elle forte, modérée ou faible ? Est-elle positive ou négative ?

2. Causalité

- Selon les données, est-il raisonnable de conclure que le **Score Moyen** cause une augmentation ou une diminution du **Prix (€)** ? Pourquoi ou pourquoi pas ?
- Identifiez d'autres variables potentielles qui pourraient influencer cette relation (par exemple, la catégorie du produit, la marque, etc.).

3. Régression Linéaire

- Proposez une équation de régression linéaire du type $y=ax+b$ où y représente le **Prix (€)** et x le **Score Moyen**.
- Interprétez le coefficient a (pente) : que représente-t-il dans ce contexte ?
- Estimez le prix d'un produit avec un **Score Moyen** de 4,2 à l'aide de l'équation de régression.

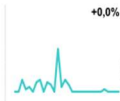
Partie II : Identification et Compréhension des KPIs

1. **KPI d'Overview** : Identifiez et expliquez brièvement les quatre indicateurs clés (KPI) présentés dans la partie "OVERVIEW" du tableau de bord (mentionner l'intitulé, la valeur et en quoi ils sont importants).
2. **Interprétation des KPI** : Pour chacun de ces indicateurs, dites quel type de performance cherche-t-on à mesurer et quelle interprétation on peut en donner.
3. **Types de projets** : Commentez les résultats présentés par le graphique à barres de type "PROJECT TYPE". Quelles conclusions peut-on en tirer?
4. **Répartition Géographique** : Expliquez l'utilité de la carte de localisation des projets, et quels types d'informations elle nous apporte.
5. **Évolution des Budgets et Dépenses** : Décrivez le graphe de type courbe dans la partie "BUDGET & EXPENSE TREND". Quelles informations pouvez-vous retirer des graphes et de la position des deux courbes (Budget et dépenses)?
6. **Niveau de priorités**: Analysez les données du graphe de type "PRIORITY". Quelles sont les priorités mis en place par le manager de ce projet, en termes de nombre de projet (et en terme de priorisation)?
7. **Santé des Projets** : En utilisant le graphe circulaire (donut) "OVERALL PROJECT HEALTH", expliquez la répartition des projets dans chaque catégorie (On Track, Needs Attention, Not Set).

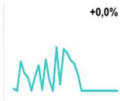


PROJECT MANAGEMENT SUMMARY

83
TOTAL PROJECTS



\$545 500
TOTAL EXPENSE



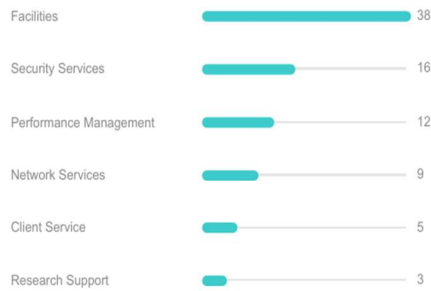
\$856 220
TOTAL BUDGET



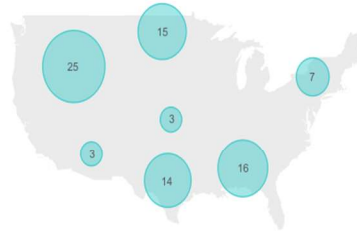
64%
BUDGET UTILISATION



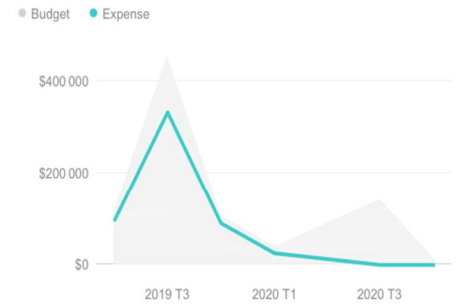
PROJECT TYPE



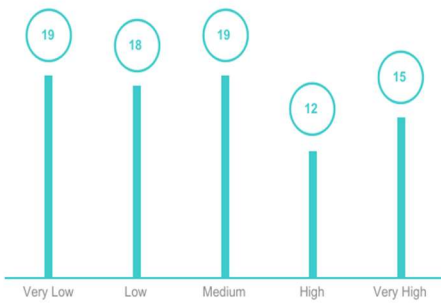
PROJECT LOCATION



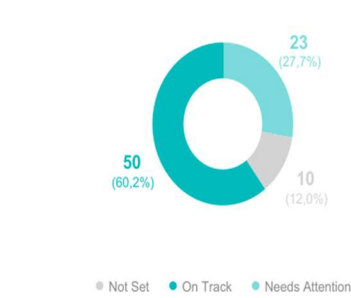
BUDGET & EXPENSE TREND



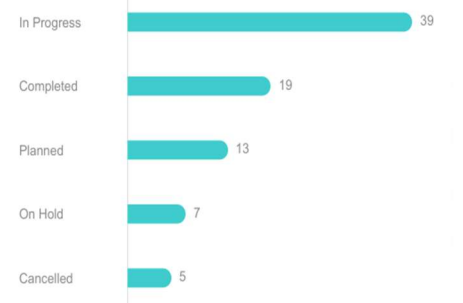
PRIORITY



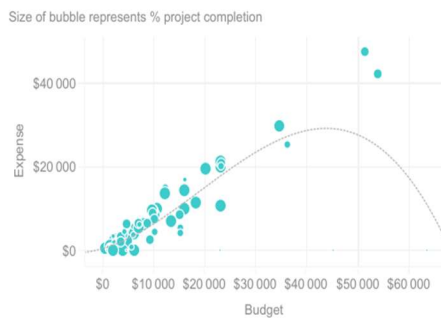
OVERALL PROJECT HEALTH



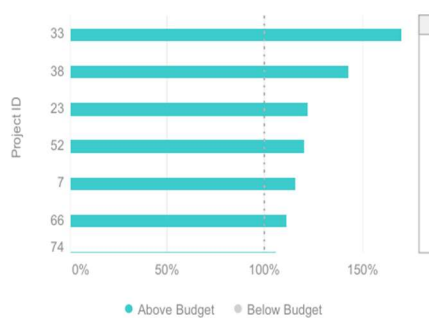
PROJECT STATUS



BUDGET VS. EXPENSES



BUDGET UTILISATION



PROJECT MANAGERS

