

Test général - Data Scientist MVola

Questions

Essayez de répondre au maximum de questions possibles

Aptitude quantitative

Instructions : choisissez la meilleure réponse parmi les cinq choix de réponse.

Question 1 :

Une famille paie 800 \$ par année pour un régime d'assurance qui paie 80 pour cent des premiers 1 000 \$ de dépenses et 100 pour cent de tous les frais médicaux par la suite. Au cours d'une année donnée, le montant total payé par la famille sera égal au montant payé par le régime lorsque les frais médicaux de la famille totaliseront.

- A. \$1,000
- B. \$1,200
- C. \$1,400
- D. \$1,800
- E. \$2,200

Question 2 :

Des sandwiches au fromage, à la Bologne et au beurre d'arachide ont été préparés pour un pique-nique dans un rapport de 5 à 7 pour 8. Si un total de 120 sandwiches étaient préparés, combien de sandwiches à la Bologne étaient-ils faits ?

- A. 15
- B. 30
- C. 38
- D. 42
- E. 48

Question 3 :

Un évier contient exactement 12 litres d'eau. Si l'eau est évacuée de l'évier jusqu'à ce qu'elle contienne exactement 6 litres d'eau de moins que la quantité évacuée, combien de litres d'eau ont été évacués ?

- A. 2
- B. 3
- C. 4.5
- D. 6
- E. 9

Question 4 :

La société C vend une gamme de 25 produits à un prix de détail moyen de 1 200 \$. Si aucun de ces produits ne se vend à moins de 420 \$ et que 10 exactement se vendent à moins de 1 000 \$, quel est le prix de vente le plus élevé possible du produit le plus cher ?

- A. \$2,600
- B. \$3,900
- C. \$7,800
- D. \$11,800
- E. \$18,200

Question 5 :

Dans un certain jeu, chaque joueur marque 2 points ou 5 points. Si n joueurs marquent 2 points et m joueurs marquent 5 points et que le nombre total de points marqués est de 50, quelle est la différence la moins positive possible entre n et m ?

- A. 1
- B. 3
- C. 5
- D. 7
- E. 9

Raisonnement

Question 1 :

À Los Angeles, un candidat politique qui achète de la publicité radiophonique à saturation obtiendra une reconnaissance maximale de son nom.

L'énoncé ci-dessus traduit logiquement lequel des énoncés suivants ?

- A. La publicité radiophonique est le facteur le plus important des campagnes politiques à Los Angeles.
- B. La reconnaissance maximale du nom à Los Angeles aidera un candidat à gagner un pourcentage plus élevé de votes exprimés dans la ville.
- C. La publicité radiophonique à saturation touche tous les secteurs démographiquement distincts de la population électorale de Los Angeles.
- D. Pour une reconnaissance maximale de nom, un candidat n'a pas besoin de dépenser sur des canaux médiatiques autres que la publicité radiophonique.
- E. Le bilan d'un candidat dans la région de Los Angeles n'aura guère d'effet sur la reconnaissance de son nom dans ce pays.

Question 2 :

Le taux de crimes violents dans cet État est en hausse de 30% par rapport à l'année dernière. La faute incombe entièrement à notre système judiciaire : Récemment, les peines de nos juges ont été si

clémentes que les criminels peuvent maintenant faire presque tout sans craindre une longue peine de prison.

L'argument ci-dessus serait affaibli s'il était vrai que :

- A. 85% des autres États du pays ont des taux de criminalité inférieurs à ceux de cet État.
- B. La criminalité en col blanc dans cet État a également augmenté de plus de 25 pour cent l'année dernière.
- C. 35 pour cent des policiers de cet État ont été licenciés l'année dernière en raison de coupes budgétaires.
- D. Les sondages montrent que 65 pour cent de la population de cet État s'oppose à la peine capitale.
- E. L'État a embauché 25 nouveaux juges l'année dernière pour compenser les décès et les départs à la retraite.

Question 3 :

L'augmentation du nombre d'articles de journaux exposés en tant que fabrications sert à renforcer l'affirmation selon laquelle les éditeurs sont plus intéressés par l'augmentation de la diffusion que par l'impression de la vérité. Même les publications mineures disposent d'un personnel pour vérifier ces fraudes évidentes.

L'argument ci-dessus suppose que :

- A. Les articles de journaux présentés comme des fabrications sont un phénomène récent.
- B. Tout ce qu'un journal imprime doit être vérifiable dans les faits.
- C. La vérification des faits est plus complète pour les publications mineures que pour les principales.
- D. Ce n'est que récemment que les journaux ont admis avoir publié des articles intentionnellement frauduleux.
- E. Les éditeurs de journaux sont les personnes qui décident quoi imprimer dans leurs journaux

Question 4 :

À maintes reprises, il a été démontré que les étudiants qui fréquentent des collèges avec de faibles ratios professeurs / étudiants reçoivent la formation la plus complète. En conséquence, lorsque mes enfants seront prêts à aller à l'université, je m'assurerai qu'ils fréquentent une école avec une très petite population étudiante.

Lequel des énoncés suivants, s'il est vrai, identifie la plus grande faille dans le raisonnement ci-dessus ?

- A. Un faible ratio professeurs / étudiants est l'effet d'une éducation bien équilibrée et non sa source.
- B. L'intelligence doit être considérée comme le résultat de l'environnement de l'enfance et non de l'enseignement supérieur.
- C. Une très petite population étudiante ne garantit pas à elle seule un faible ratio professeurs / étudiants.
- D. Les désirs et préférences des parents déterminent rarement le choix d'un enfant d'un collège ou d'une université.

E. Les étudiants doivent profiter du faible ratio professeurs / étudiants en choisissant intentionnellement de petites classes.

Question 5 :

Tous les philosophes allemands, à l'exception de Marx, sont des idéalistes.

De laquelle des affirmations suivantes peut-on déduire le plus correctement l'énoncé ci-dessus ?

- A. À l'exception de Marx, si quelqu'un est un philosophe idéaliste, alors il ou elle est allemande.
- B. Marx est le seul philosophe non allemand qui soit idéaliste.
- C. Si un Allemand est un idéaliste, alors il ou elle est un philosophe, tant qu'il ou elle n'est pas Marx.
- D. Marx n'est pas un philosophe allemand idéaliste.
- E. En dehors du philosophe Marx, si quelqu'un est un philosophe allemand, alors il ou elle est un idéaliste.

Data understanding

Instructions : Dans chacun des problèmes, une question est suivie de deux énoncés contenant certaines données. Vous devez déterminer si les données fournies par les déclarations sont suffisantes pour répondre à la question.

Choix de réponse :

- A. si l'énoncé (1) suffit à lui seul pour répondre à la question, mais l'énoncé (2) en lui-même ne l'est pas
- B. si la déclaration (2) en elle-même suffit à répondre à la question, mais la déclaration (1) en elle-même ne l'est pas
- C. si les énoncés (1) et (2) pris ensemble suffisent à répondre à la question, même si aucun des énoncés ne suffit en soi
- D. si l'une ou l'autre des déclarations suffit à elle seule pour répondre à la question
- E. si les énoncés (1) et (2) pris ensemble ne sont pas suffisants pour répondre à la question, nécessitant davantage de données relatives au problème

Question 1 :

$x = y$?

(1) $x^2 - y^2 = 0$

(2) $(x - y)^2 = 0$

Question 2 :

Si R est un entier, R est-il divisible par 3 ?

(1) $2R$ est divisible par 3

(2) $3R$ est divisible par 3

Question 3 :

S'il ne s'est pas arrêté en cours de route, quelle vitesse Bill a-t-il en moyenne sur son trajet de 3 heures ?

(1) Il a parcouru un total de 120 milles.

(2) Il a parcouru la moitié de la distance à 30 milles à l'heure et la moitié à 60 milles à l'heure.

Question 4 :

Est-ce que $x + y$ est positif ?

(1) $x - y$ est positif.

(2) $y - x$ est négatif.

Question 5 :

Un acheteur a acheté une cravate et une ceinture lors d'une vente. Quel article a-t-il acheté à la plus grande valeur monétaire ?

(1) Il a acheté la cravate à un rabais de 20%.

(2) Il a acheté la ceinture à un rabais de 25%

Aptitude Story Telling et restitution

« SO WHAT ? » TEST

Instructions :

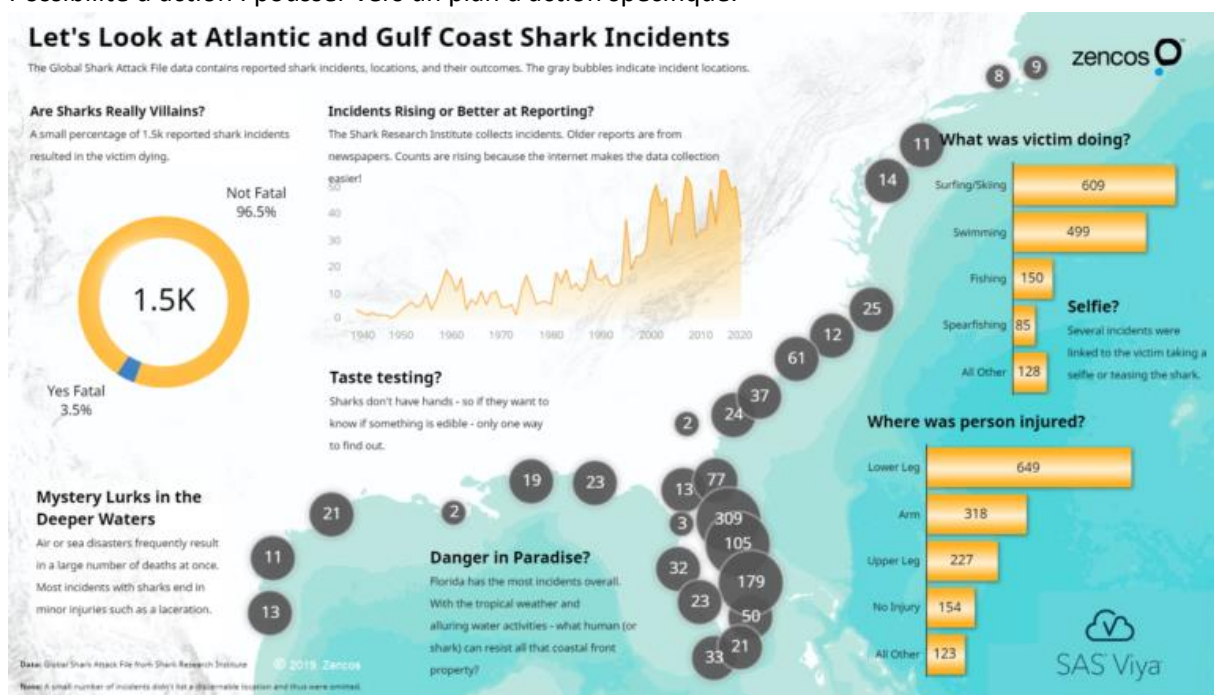
Resituez ce que vous pouvez en tirer du graphe présenté. Vous pouvez vous appuyer sur les points suivants :

Compréhension : communiquer sur le contenu du graphe

Explicabilité : relier la visualisation aux phénomènes mesurés (par exemple, que se passe-t-il ?)

Affect : comprendre les conclusions intellectuellement.

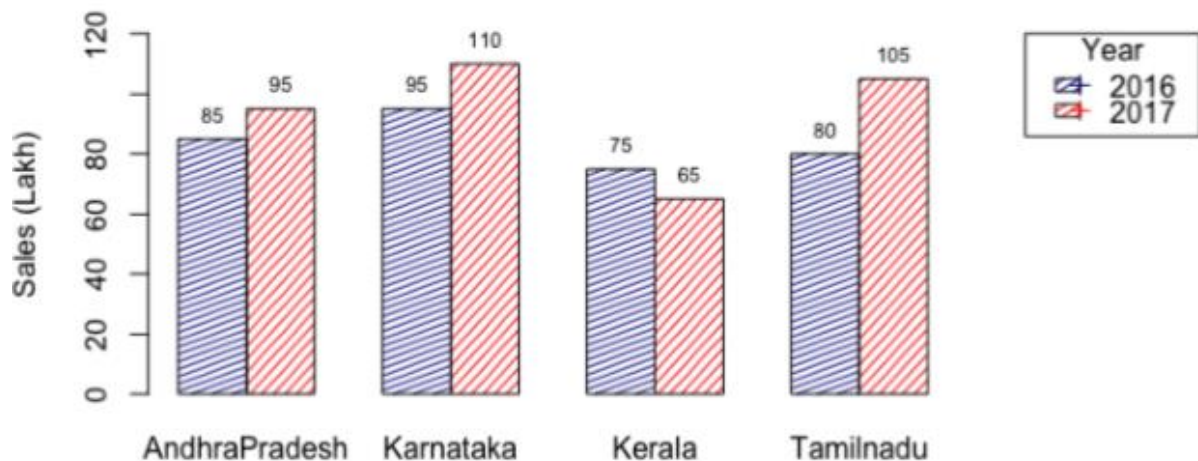
Possibilité d'action : pousser vers un plan d'action spécifique.



<https://www.zencos.com/blog/infographics-basics-example-design-guide/>

INTERPRETATION GRAPHIQUE

Le graphique à barres ci-dessous montre les ventes de livres de non-fiction d'un magasin en ligne, pendant deux années consécutives 2016 et 2017, pour quatre états de l'Inde.



Lesquelles des affirmations suivantes sont correctes ?

- A. Le rapport des ventes totales du Karnataka pour les deux années aux ventes totales du TamilNadu pour les deux années est de 41:37
- B. Le total des ventes pour l'année 2017 est de Rs. 335
- C. Le chiffre d'affaires moyen en 2016 est de 89,33% du chiffre d'affaires moyen de 2017.
- D. Les ventes moyennes dans le sud de l'Inde sur deux ans sont de Rs 355
- E. The sales of Kerala is dropped by 13%
- F. The sales of TamilNadu has increased by 25%.
- G. The growth in book sales in Andhra Pradesh is larger than the Karnataka.
- H. (d) Overall growth in book sales is 11.9%

Probabilité & Statistique

Question 1 :

Soa a 2 enfants. Soit

- A l'événement où l'aîné est une fille,
- B l'événement qu'au moins un enfant est une fille
- et C l'événement que les deux enfants sont des filles.

Lesquelles des affirmations suivantes sont correctes ?

- A. $P(C|A) = \frac{1}{2}$
- B. $P(C|B) = \frac{1}{2}$
- C. $P(C|A) = \frac{1}{3}$
- D. $P(C|B) = \frac{1}{3}$

Choisissez la réponse la plus probable

Question 2 :

Lors de l'analyse des données catégorielles, le graphique suivant n'est pas approprié

- A. Pie chart

- B. Diagramme de Pareto.
- C. Stem and leaf display
- D. Diagramme à barres.
- E. ils sont tous appropriés.

Question 3 :

Une mesure calculée uniquement à partir d'un échantillon de la population est appelée

- A. Un paramètre.
- B. Un recensement.
- C. Une statistique.
- D. La méthode scientifique.
- E. Tout ce qui précède

Question 4 :

Une mesure calculée uniquement à partir d'un échantillon de la population est appelée

- A. Un paramètre.
- B. Un recensement.
- C. Une statistique.
- D. La méthode scientifique.
- E. Tout ce qui précède

Pour les questions 5 à 7 considérez l'exemple suivant :

Le temps d'attente moyen dans la section Shop d'une chaîne populaire est de 4 minutes avec un écart type de 1 minute.

Question 5 :

Quelle est la probabilité qu'un client doive attendre plus de 3,5 minutes ?

- A. 0,5000
- B. 0,2375
- C. 0,7523
- D. 0,3085
- E. 0,6915

Question 6 :

La même chaîne interroge plus tard 1000 clients pour voir si elle devrait ajouter un autre service à son Shop. La chaîne compte 50 Shops mais, en supposant que les clients de chaque shop ont des comportements similaires, ne choisit au hasard que 10 des magasins pour participer à l'enquête. Chaque magasin échantillonne ensuite 100 clients. Quelle méthode d'échantillonnage la chaîne de Shop a-t-elle utilisée ?

- A. Échantillonnage en cluster
- B. Échantillonnage non probabiliste

- C. Échantillonnage aléatoire simple
- D. Échantillonnage stratifié
- E. Échantillonnage systématique

Question 7 : Case study

Les propriétaires de shop sont préoccupés par la baisse de leurs ventes au cours des derniers mois. En conséquence, ils veulent déterminer s'il existe une différence dans la motivation de l'animateur de vente rémunéré au taux horaire majoré de la commission ou uniquement à la commission (mais à un taux de pourcentage plus élevé). Sur 24 employés nouvellement sélectionnés au hasard, 12 ont été rémunérés au taux horaire et 12 à la commission uniquement. Les données suivantes représentent les ventes en volume réalisées au cours du premier mois de travail.

Taux horaire	Commission
147	330
224	472
118	195
209	489
126	386
372	462
197	509
260	312
372	227
451	325
447	518
328	476

Résultats du test d'hypothèse

	Taux horaire	Commission
Mean	270.9166667	391.75
Variance	14407.90152	12557.47727
Observations	12	12
Pooled Variance	13482.68939	
Hypothesized Mean Difference	0	
Df	22	
t Stat	-2.549025127	
P(T<=t) one-tail	0.009146094	
t Critical one-tail	1.717144187	
P(T<=t) two-tail	0.018292187	
t Critical two-tail	2.073875294	

Pour déterminer s'il y a une différence dans le volume des ventes des deux méthodes de motivation, ils ont décidé d'analyser les données avec un test d'hypothèse à un niveau de signification de 5%.

1. Etablir l'hypothèse nulle : phrase et expression statistique

2. Etablir l'hypothèse alternative : phrase et expression statistique
3. Construire le test d'hypothèse et conclure par rapport au problème posé
4. Dans cette situation, décrire ce qu'est l'erreur type 1

Modélisation

Les données suivantes représentent le nombre de résidents sans emploi de Brokenleggen au cours des 20 trimestres économiques de 1996 à 2000.

Brokenleggen est un village médiéval des Alpes suisses dont le taux de chômage est principalement influencé par l'afflux de skieurs, en particulier pendant les mois enneigés d'octobre à mars.

Quarter/Year	Sans emploi	Quarter/Year	Sans emploi
Mars 1996	105	Septembre 1998	241
Juin 1996	214	Décembre 1998	128
Septembre 1996	226	Mars 1999	121
Décembre 1996	126	Juin 1999	217
Mars 1997	112	Septembre 1999	250
Juin 1997	232	Décembre 1999	122
Septembre 1997	224	Mars 2000	126
Décembre 1997	131	Juin 2000	229
Mars 1998	106	Septembre 2000	259
Juin 1998	208	Décembre 2000	133

Commenter sur la saisonnalité du nombre de résidents sans emploi de Brokenleggen (Tips : Tracer le graphique de la série temporelle, calculer l'indice de saisonnalité pour chaque trimestre, estimer en utilisant une simple régression linéaire)

Prédire le nombre de résidents sans emploi pour Décembre 2001 avec une méthode que vous avez mis en place

Programming

Les instructions sont déjà disponibles dans les dossiers FindXWithRegex et HowmanyRight

SQL Test

Répondez aux questions dans SqlTestMVola.pdf

NB

Format : Vous êtes libre d'envoyer vos résultats directement par mail ou/et nous envoyer le lien github publique contenant les résultats

Ce sujet reste confidentiel, ne pas diffuser ou partager s'il vous plaît.

Contact : nomena.nyhoavy@mvola.mg