

# Table des matières

## RESSOURCES

XIV

## 1 Notebooks - Python - Pandas 1

<b>1</b>	Python . . . . .	<b>2</b>
<b>A</b>	Jupyter Notebook . . . . .	2
<b>1</b>	Présentation . . . . .	2
<b>2</b>	Installation . . . . .	2
<b>3</b>	Anaconda . . . . .	2
<b>4</b>	Nouveau notebook . . . . .	4
<b>5</b>	Manipulations complémentaires . . . . .	5
<b>B</b>	Rappels de base . . . . .	6
<b>1</b>	Opérations élémentaires . . . . .	6
<b>2</b>	Variable : type, affichage . . . . .	6
<b>3</b>	Boucles . . . . .	7
<b>C</b>	Données structurées . . . . .	7
<b>1</b>	Listes . . . . .	7
<b>2</b>	Dictionnaires . . . . .	9
<b>D</b>	Modules et fonctions . . . . .	9
<b>1</b>	Module math . . . . .	9
<b>2</b>	Module pylab . . . . .	10
<b>3</b>	Module random . . . . .	12
<b>4</b>	Module numpy . . . . .	12
<b>5</b>	Module scipy.stats . . . . .	13
<b>6</b>	Fonctions . . . . .	13
<b>E</b>	Modules plotly, seaborn, geopandas, networkx . . . . .	15
<b>1</b>	Installations . . . . .	16
<b>2</b>	Ouvrir Jupyter Notebook dans un autre navigateur . . . . .	17
<b>2</b>	Traitement des données . . . . .	<b>18</b>
<b>A</b>	Fichiers tabulaires de données . . . . .	18

<b>B</b>	Lecture, sélection, croisement, regroupement de données . . . . .	18
<b>1</b>	Le module pandas . . . . .	18
<b>2</b>	Lecture avec pandas . . . . .	19
<b>3</b>	Sélection de données . . . . .	20
<b>4</b>	Croisement de données . . . . .	20
<b>5</b>	Data visualisation des fréquences conditionnelles & diagrammes circulaires	21
<b>6</b>	Regroupement de données . . . . .	23
<b>7</b>	Data visualisation des indicateurs statistiques & la boîte à moustaches .	24
<b>8</b>	Mémento pandas . . . . .	28
<b>Test</b>	QCM . . . . .	31

## A Premières Data visualisations

### 2 Les historiques 37

<b>1</b>	Carte statistique avec python . . . . .	38
<b>A</b>	Représentation cartographique de données statistiques . . . . .	38
<b>1</b>	Carte statistique . . . . .	38
<b>2</b>	Carte statique -vs- dynamique . . . . .	38
<b>3</b>	OpenStreetMap . . . . .	38
<b>4</b>	Mapbox . . . . .	39
<b>B</b>	Le module plotly . . . . .	39
<b>1</b>	Obtention d'une clef pour utiliser des cartes Mapbox . . . . .	39
<b>2</b>	plotly.graph_objects . . . . .	39
<b>3</b>	plotly.express . . . . .	42
<b>4</b>	Carte de chaleur . . . . .	44
<b>2</b>	John Snow et le choléra . . . . .	46
<b>3</b>	Carte de Charles-Joseph Minard : Campagne de Russie . . . . .	50
<b>4</b>	Espérance de vie en France « façon Minard » . . . . .	55
<b>Problème 1</b>	Roses de Nightingale. . . . .	58

### 3 Puissance de la data visualisation 67

<b>1</b>	Cartes choroplèthes sur une mappemonde . . . . .	68
<b>A</b>	Coloration de plusieurs pays . . . . .	68
<b>1</b>	Exemples de coloration avec des codes ISO-3 . . . . .	68
<b>2</b>	Exemple de coloration avec des noms de pays . . . . .	71
<b>B</b>	L'espérance de vie en 2015 dans le monde . . . . .	71
<b>C</b>	La température moyenne dans le monde . . . . .	73

<b>2</b>	Carte française des mouvements de population liés à la covid . . . . .	<b>75</b>
<b>3</b>	Data visualisation des volcans dans le monde . . . . .	<b>81</b>
<b>Test</b>	Parité des femmes élues aux municipales . . . . .	<b>85</b>
<b>Problème 2</b>	Cartographie des 100 plus grandes villes des États-Unis . . . . .	<b>89</b>
<b>Problème 3</b>	Data visualisation de l'influence de 5 puissances maritimes d'Europe . . . . .	<b>96</b>
<b>Problème 4</b>	Les arbres de Paris à fort potentiel allergisant . . . . .	<b>109</b>
<b>Problème 5</b>	Data visualisations des ouragans . . . . .	<b>120</b>
<b>Problème 6</b>	Data visualisations des séismes . . . . .	<b>137</b>
<b>Problème 7</b>	Premier générateur de nombres pseudo-aléatoires d'IBM . . . . .	<b>146</b>

## B Lanceur d'alerte

<b>4</b>	<b>Usage des data visualisations</b>	<b>155</b>
<b>Problème 8</b>	Le Benin, déclaration d'âges . . . . .	<b>156</b>
<b>5</b>	<b>Usage de simulations</b>	<b>161</b>
<b>Problème 9</b>	Benford et suspicion de fraudes aux présidentielles des USA en 2016? . . . . .	<b>162</b>
<b>Problème 10</b>	Lanceur d'alerte et jeux de grattages . . . . .	<b>170</b>
<b>Problème 11</b>	Détection de fraudes au tennis corruption et mafia . . . . .	<b>176</b>

## C Le citoyen et ses finances

<b>Problème 12</b>	Concurrence Airbnb -VS- Hôtels? Un vrai data challenge . . . . .	<b>189</b>
<b>Problème 13</b>	Répartition des revenus imposables en France, des inégalités?. . . . .	<b>204</b>
<b>Problème 14</b>	Salaires des joueurs de la FIFA, comparaison des meilleurs clubs . . . . .	<b>218</b>
<b>Problème 15</b>	Le bitcoin versus un placement stable . . . . .	<b>230</b>

## D Modéliser et simuler

<b>6</b>	<b>Catastrophes ou phénomènes naturels</b>	<b>239</b>
<b>Problème 16</b>	Temps d'attente entre 2 accidents d'avions . . . . .	<b>240</b>
<b>Problème 17</b>	Crue centennale de la Seine. . . . .	<b>247</b>
<b>Problème 18</b>	Propagation d'un incendie & la percolation. . . . .	<b>255</b>
<b>Problème 19</b>	Diffusion d'un gaz & urnes d'Ehrenfest . . . . .	<b>267</b>
<b>Problème 20</b>	4 ouragans de force 5. . . . .	<b>279</b>
<b>7</b>	<b>Epidémiologie</b>	<b>287</b>

<b>Problème 21</b>	Variale vue par Bernoulli statistique & équation différentielle . . . . .	<b>288</b>
<b>Problème 22</b>	Épidémie vue par Reed & Frost - modèle du Pile ou Face . . . . .	<b>299</b>

## **8** Au football **309**

<b>Problème 23</b>	Prédire comme un bookmaker & Ligue 1 . . . . .	<b>310</b>
<b>Problème 24</b>	Scores en Premier League . . . . .	<b>324</b>
<b>Problème 25</b>	Buts à domicile vs l'extérieur . . . . .	<b>329</b>

## **E Prêt pour la photo?**

<b>1</b>	Qu'est-ce qu'une image? . . . . .	<b>343</b>
<b>2</b>	Avec python . . . . .	<b>344</b>
<b>Problème 26</b>	Une spirale naturelle. . . . .	<b>345</b>
<b>Problème 27</b>	Recherche du meilleur profil . . . . .	<b>351</b>
<b>Problème 28</b>	La transformation du photomaton. . . . .	<b>368</b>

## **F Parcourir le monde**

### **9** Carte de navigation révolutionnaire **379**

<b>Problème 29</b>	Naviguer à cap constant avec Mercator & histoire des mathématiques . . . . .	<b>380</b>
--------------------	--	------------

### **10** Plus court chemin dans un réseau réel **397**

<b>1</b>	Introduction aux graphes. . . . .	<b>398</b>
<b>2</b>	Algorithme de Dijkstra : plus court chemin. . . . .	<b>398</b>
<b>A</b>	Algorithme . . . . .	<b>398</b>
<b>B</b>	Exemple . . . . .	<b>400</b>
<b>C</b>	Utilisation du module networkx . . . . .	<b>403</b>
<b>1</b>	Première façon de définir le graphe G . . . . .	<b>403</b>
<b>2</b>	Deuxième façon de définir le graphe G . . . . .	<b>404</b>
<b>3</b>	Sommets et poids . . . . .	<b>404</b>
<b>4</b>	Dijkstra . . . . .	<b>404</b>
<b>5</b>	Autres actions . . . . .	<b>405</b>
<b>D</b>	Exemple châteaux de la Loire . . . . .	<b>405</b>
<b>Problème 30</b>	Le métro Londonien . . . . .	<b>407</b>
<b>Problème 31</b>	Aéroports du monde entier . . . . .	<b>418</b>

## G Prédire - Intelligence Artificielle

<b>11</b>	<b>La régression linéaire</b>	<b>429</b>
<b>1</b>	Moyenne & variance & écart-type . . . . .	430
<b>A</b>	Moyenne . . . . .	430
<b>B</b>	Variance . . . . .	431
<b>C</b>	Ecart-type . . . . .	432
<b>D</b>	Normalisation d'une variable $X$ . . . . .	433
<b>2</b>	Covariance et corrélation d'un couple $(X, Y)$ . . . . .	434
<b>A</b>	Nuage de points & point moyen . . . . .	434
<b>B</b>	Covariance . . . . .	435
<b>C</b>	Covariance et corrélation . . . . .	438
<b>Problème 32</b>	Les criquets comme thermomètre? . . . . .	440
<b>4</b>	Régression linéaire . . . . .	450
<b>A</b>	Pertinence de trouver $a$ et $b$ tel que $Y \approx aX + b$ . . . . .	450
<b>1</b>	Cas 1 : le nuage est formé de points alignés . . . . .	450
<b>2</b>	Cas 2 : le nuage a une allure rectiligne . . . . .	451
<b>B</b>	Galton et la régression . . . . .	454
<b>C</b>	Mélanome malin aux USA . . . . .	455
<b>Problème 33</b>	Les allergisants de retour à Paris, à quoi ressemblent-ils? . . . . .	458
<b>12</b>	<b>Machine learning</b>	<b>463</b>
<b>Problème 34</b>	Les iris, "à partir de tes dimensions je te dirai qui tu es". . . . .	464
<b>Problème 35</b>	Détection du diabète . . . . .	479
<b>Problème 36</b>	Prédire le dépassement d'un seuil de concentration d'ozone . . . . .	493
	<b>INDEX</b>	<b>519</b>
	<b>CRÉDITS PHOTOS</b>	<b>522</b>