**Cahier des charges (CDC)**

**Analyse de données**

Auteurs :

Tanguy Franky

Boucly Kévin

Plessis Loïc

Le 18/03/22

Sommaire

**1) Introduction**  3

1.1 - Motivations. 3

1.2 - Contexte 3

**2) Documentation**  4

2.1 - Terminologie métier 4

2.2 - Bibliographie succincte. 5

***3) Profil des utilisateurs finaux***  5

3.1 - En utilisation. 5

3.2 - En exploitation. 5

3.3 - En maintenance. 5

3.4 - Environnement d’utilisation. 5

**4) Fonctions à réaliser**  6

4.1 - Ce que le système doit faire. 6

4.2 - Diagrammes des données 7-10

4.3 Vue et procédure stockée

**5) Aspects contractuels**  10

5.1 - Délais et protocole de livraison. 11

5.2 - Aspect juridique et commercial. 12

**1) Introduction**

Vous êtes employés par le CSOB (Central Statistical Office of Breizhmeiz). Vous venez de recevoir un dataset à analyser et votre chef de bureau veut une première expertise sous forme de notebooks dans 2 jours et un cahier des charges à livrer au client.

**1.1 Motivations**

**1.1.1 Le client**

Ubisoft est une entreprise française de développement, d'édition et de distribution de jeux vidéo.

**1.1.2 Le problème**

Livrer une première expertise du Dataset “ Ventes Jeux vidéo” sous forme de notebooks et un cahier des charges à livrer au chef de bureau qui sera envoyé ensuite au client pour qu’il puisse exploiter les données analysées.

**1.1.3 Les objectifs**

* Réaliser le MCD de la base correspondante
* Mettre en place la base de données
* Charger le dataset dans la base à l'aide d'un script Python
* Créer une vue et une procédure stockée
* Rédiger un Notebook contenant le MCD, une description du travail de création des bases et le code SQL de la vue et de la procédure stockée
* Analyser les données en Python sur un Notebook
* Visualiser les données via les bibliothèques Python adaptées sur le même Notebook

**2) Documentation**

**2.1 Terminologie métier**

**MCD : modèle conceptuel de données :** a pour but d'écrire de façon formelle les données qui seront utilisées par le système d'information. Il s'agit donc d'une représentation des données, facilement compréhensible, permettant de décrire le système d'information à l'aide d'entités.

**SQL**: Le langage **SQL** (Structured Query Language) est un langage informatique utilisé pour exploiter des bases de données.

**Python:** Python est un **langage de programmation interprété**, multi-paradigme et multiplateformes. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet.

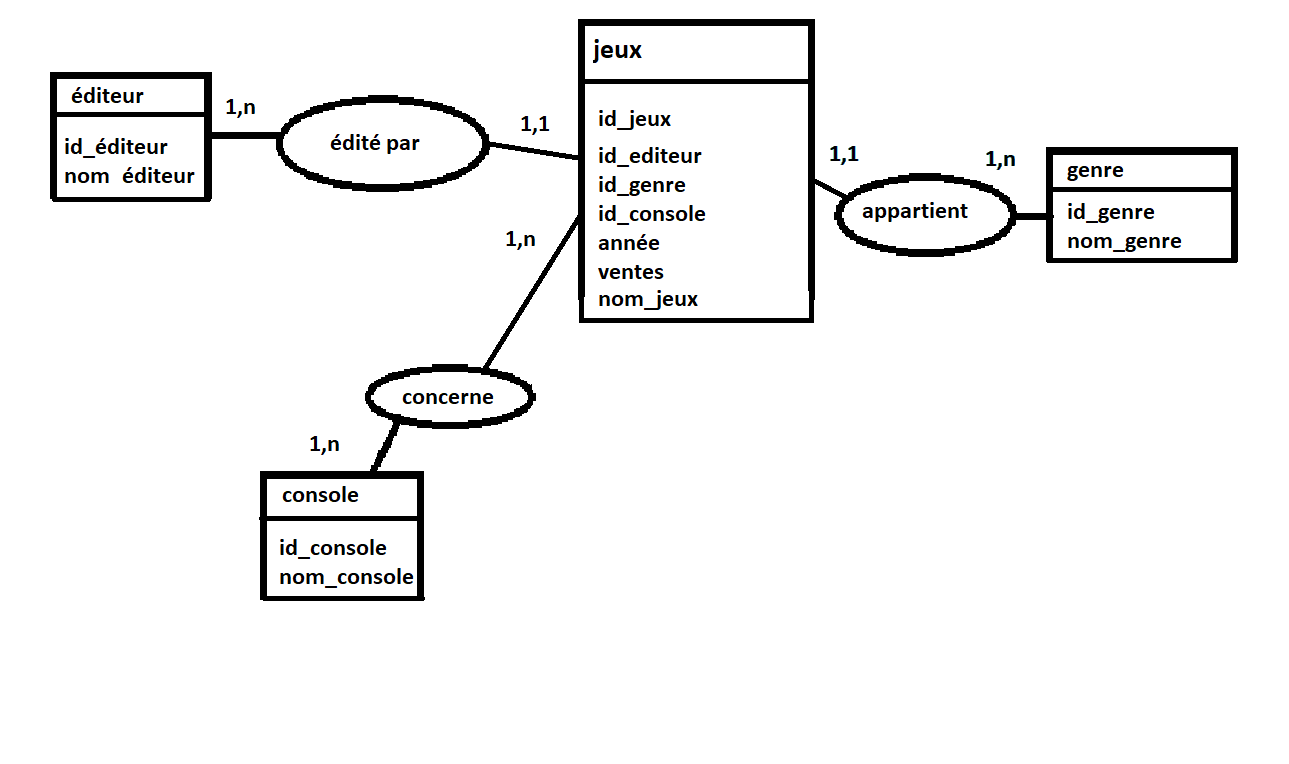
**Notebook:** **Jupyter Notebook** permet aux développeurs de partager du code et de l'exécuter dans la même interface utilisateur. Il peut associer du code, des graphiques, des visualisations et du texte dans des **notebooks** - ou cahiers

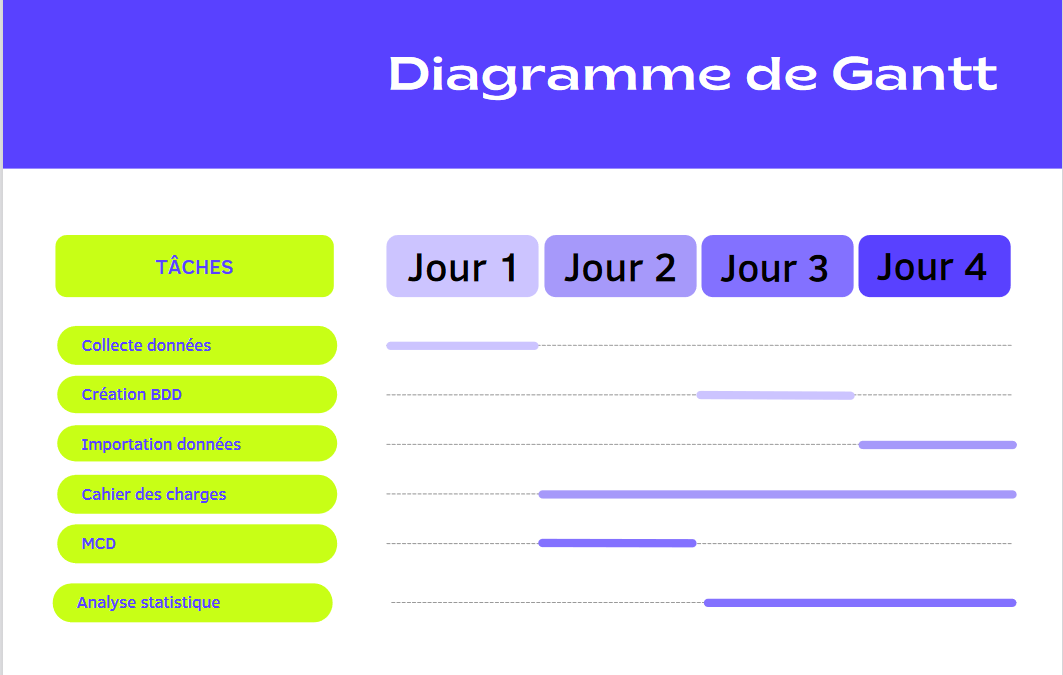
**Vue:** Une **vue** est une représentation logique d'une requête utilisateur réalisée sur une ou plusieurs tables. C'est une table virtuelle basée sur les données des tables de la base. Elle se comporte ensuite comme si elle était une table du moins pour ce qui est de son interrogation.

**Procédure stockée**: une **procédure stockée** (ou stored **procedure** en anglais) est un ensemble d'instructions **SQL** précompilées, **stockées** dans une base de données et exécutées sur demande par le SGBD qui manipule la base de données.

**2.2 Bibliographie succincte**

**Schéma modèle conceptuel des données:**





**3) Profil des utilisateurs finaux**

Professionnels dans l’édition de jeux vidéo

**3.1 En utilisation**

Développeurs, Data scientist interne à l'entreprise, éditeurs de jeux vidéo

**3.2 En exploitation**

Les utilisateurs finaux ont des connaissances en programmation et les data scientist savent analyser les données.

**3.3 En maintenance**

Pas de maintenance, il s’agit d’une analyse.

**3.4 Environnement d’utilisation**

Windows 10

**4 ) Fonctions à réaliser**

Analyse sur un dataset nommé “ Ventes Jeux vidéo”

**4.1 Ce que le système doit faire**

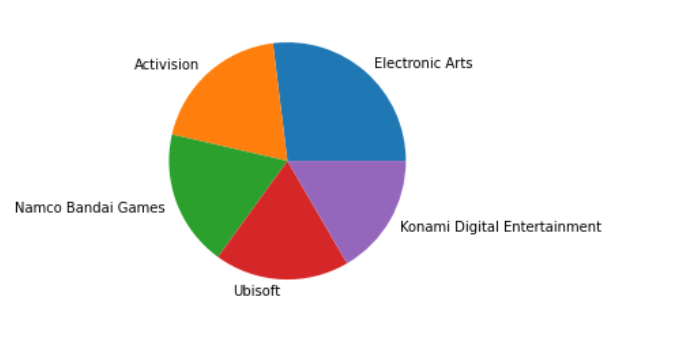
Présentation , Modèle conceptuel (MCD) et Analyse statistiques de la base de données (Dataset) “ Ventes Jeux vidéo”.

L’Analyse statistique comprend :

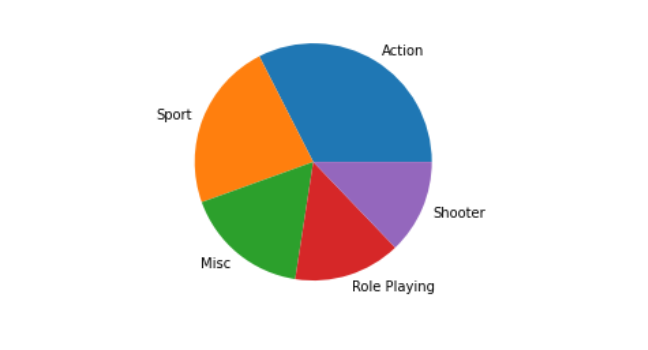
* le nombre de plateforme par jeux
* le nombre de jeux par plateforme
* le nombre de jeux par catégorie
* le nombre de jeux par éditeur
* la répartition du nombre de jeux publié par éditeur
* répartition des jeux par genre
* le nombre de jeux sortis par consoles

**4.2 Diagrammes des données**

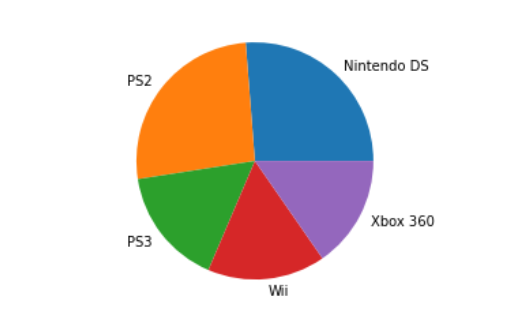
**Sélection graphique des données les plus pertinentes ( année chiffres : 2014)**



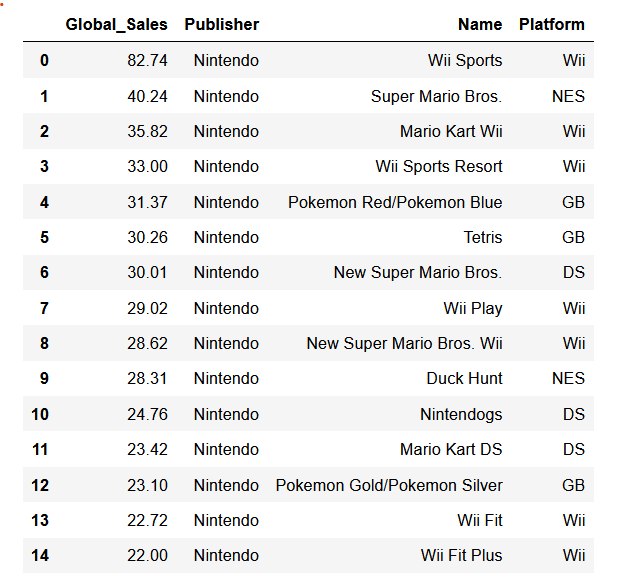
Répartition du nombre de jeux publiés par éditeurs



Répartition du nombre de jeux publiés par genre

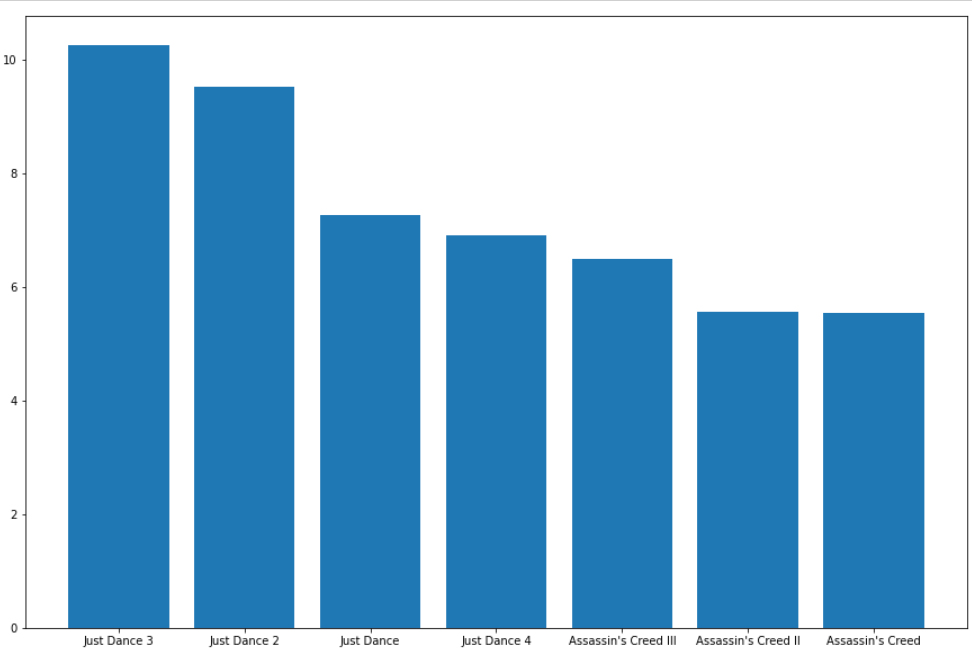


Répartition du nombre de jeux publiés par console



On remarque que l’éditeur qui a le plus vendu de jeux vidéo est Nintendo

Concernant Ubisoft:



On remarque que Just Dance et Assasin's Creed sont les deux licences les plus vendues d'Ubisoft.( chiffres graphique en millions de ventes)

**4.3 Vue et procédure stockée**

**Exemple d’une vue qui permet d’afficher tous les jeux vendus à plus de 20 millions d’exemplaires:**

requete\_vue='''CREATE VIEW Best\_Games AS # Création d'une vue

SELECT ventes, nom\_jeu, annee

FROM jeux

WHERE ventes > 20 ;

SELECT \* FROM Best\_Games;'''

Exemple d’une procédure stocké qui retourne tous les jeux sortis dans l’année qu’on lui passe en paramètre :

procedure\_stockee = '''DELIMITER //

CREATE PROCEDURE proc

(IN var CHAR(50))

BEGIN

SELECT \*

FROM jeux

WHERE année =var;

END //

DELIMITER ;

**5) Aspects contractuels**

**5.1 - Délais et protocole de livraison.**

**Date de livraison** : 01/04/2022

**Livrables :**

**- le fichier SQL contenant la Base de données**

**- Un notebook contenant :**

**- La présentation de la base (qui contient le MCD du dataset)**

**- Un script python contenant le code de l'exploitation de la base**

**- Une analyse statistique**

**- Un cahier des charges**

**5.2 - Aspect juridique et commerciaux**

***Fait le 18/03/2022, A Brest***

**Le représentant de Simplon Le représentant de CSOB Signature Signature**

***(Précédé de la mention « lu et approuvé ») (Précédé de la mention « lu et approuvé »)***