**Rapport de Projet :** Analyse des Retards des Trains SNCF en Relation avec le Trafic Routier et les Conditions Météorologiques

**Introduction**

Ce projet vise à analyser les retards des trains SNCF en relation avec le trafic routier et les conditions météorologiques en Île-de-France. En combinant des données provenant de plusieurs sources, nous cherchons à identifier les facteurs clés qui influencent les retards des trains, et à proposer des visualisations claires et informatives pour mieux comprendre ces relations.

**Objectifs**

1. **Ingestion des Données :** Collecter et centraliser les données provenant de différentes sources.
2. **Transformation des Données :** Nettoyer et transformer les données pour les rendre exploitables.
3. **Visualisation :** Créer des visualisations interactives pour présenter les résultats de l'analyse et effectuer les jointures nécessaires dans Power BI.

**Sources de Données**

1. **Données de Régularité des Trains (SNCF) :**
   * **Fichier :** regularite-mensuelle-tgv-aqst.xlsx
   * **Contenu :** Informations sur les retards des trains TGV en France, incluant les dates, gares, types de train, et durées de retard.
2. **Données de Trafic Routier :**
   * **Fichier :** observatoire-parisien-des-mobilites-barometre-trimestriel-des-deplacements.csv
   * **Contenu :** Informations sur le trafic routier à Paris, incluant les dates, lieux, nombre de véhicules, vitesses moyennes, et taux de congestion.
3. **Données Météorologiques :**
   * **Fichier :** meteo-france.csv
   * **Contenu :** Informations météorologiques, incluant les dates, températures, précipitations, et conditions météorologiques générales.

Shema :

Une image contenant texte, capture d’écran, diagramme, conception

Description générée automatiquement

**Étapes du Projet**

**1. Ingestion des Données**

* **Création des Conteneurs dans Azure Data Lake Storage Gen2 :**
  + raw-data : Stockage des données brutes.
  + processed-data : Zone pour les données transformées.
* **Pipelines d'Ingestion :**
  + **Ingestion des Données SNCF :**
    - Pipeline : Ingestion\_SNCF\_To\_Processed
    - Activité : Copy\_SNCF\_To\_Processed
    - Source : raw-data/regularite-mensuelle-tgv-aqst.xlsx
    - Destination : processed-data/regularite-mensuelle-tgv-aqst.xlsx
  + **Ingestion des Données de Trafic Routier :**
    - Pipeline : Ingestion\_Traffic\_To\_Processed
    - Activité : Copy\_Traffic\_To\_Processed
    - Source : raw-data/observatoire-parisien-des-mobilites-barometre-trimestriel-des-deplacements.csv
    - Destination : processed-data/observatoire-parisien-des-mobilites.csv
  + **Ingestion des Données Météorologiques :**
    - Pipeline : Ingestion\_Weather\_To\_Processed
    - Activité : Copy\_Weather\_To\_Processed
    - Source : raw-data/meteo-france.csv
    - Destination : processed-data/meteo-france.csv

**2. Transformation des Données**

* **Transformation des Données SNCF :**
  + **Pipeline :** Transform\_SNCF\_Data
  + **Activité :** Copy\_SNCF\_With\_Transformations
  + **Transformations :**
    - Sélectionner les colonnes pertinentes (date, gare, type\_de\_train, duree\_du\_retard)
    - Filtrer les données incorrectes
  + **Destination :** processed-data/transform\_snc\_data.parquet

**3. Visualisation et Jointures dans Power BI**

* **Connecter Power BI à Azure Data Lake Storage Gen2 :**
  + Ouvrir Power BI Desktop.
  + Cliquer sur "Obtenir des données" et sélectionner "Azure blob Storage ".
  + Connecter à votre compte de stockage et sélectionner les fichiers dans le conteneur processed-data.

Une image contenant texte, capture d’écran, logiciel, nombre

Description générée automatiquement

Une image contenant texte, capture d’écran, Police, diagramme

Description générée automatiquement

**Analyse Complète des Données sur la Régularité des Trains TGV (2022-2024)**

**Introduction**

Les rapports "Retard des trains par gare" pour les années 2022, 2023 et 2024 mettent en lumière les performances des trains en termes de ponctualité, de retards et d'annulations. Cette analyse se penche sur les données clés pour diverses gares ainsi que sur la performance globale des services ferroviaires nationaux et internationaux.

**Données Générales**

Sur les 1,11M de trains prévus, 154,61K ont été retardés à l'arrivée, 24,83K ont été annulés et 299,22K ont subi des retards au départ. Ces chiffres indiquent des défis significatifs en matière de ponctualité dans le réseau ferroviaire.

**Performances par Gare**

* **Paris Montparnasse**:
  + Annulations: 5,14K
  + Retards à l'arrivée: 21,31K
  + Retards au départ: 24,4K

Paris Montparnasse se distingue comme la gare avec le plus grand nombre de retards et d'annulations dans les trois catégories mentionnées.

* **Paris Lyon**:
  + Annulations: 4,24K
  + Retards à l'arrivée: 25,43K
  + Retards au départ: 28,08K

Paris Lyon suit de près avec des chiffres légèrement plus élevés pour les retards à l'arrivée et au départ, mais moins d'annulations par rapport à Paris Montparnasse.

* **Madrid**:
  + Annulations: 10
  + Retards à l'arrivée: 42
  + Retards au départ: 0

Madrid présente un contraste frappant avec un nombre négligeable de retards et d'annulations, indiquant une performance exemplaire en termes de ponctualité.

**Corrélation entre Annulations et Retards**

Le rapport sur les retards indique une corrélation significative entre le nombre de trains annulés et le nombre total d'arrivées retardées, suggérant que les annulations sont souvent prévisibles dans les contextes de nombreux retards.

**Performance des Services Ferroviaires**

Les données de la page "Analyse de la performance des services ferroviaires" montrent 96,66K annulations sur un total de 842,47K départs retardés et 346,24K arrivées retardées. La durée moyenne des trajets est de 17K minutes, soulignant les inefficacités actuelles.

**Étude de Profil des Services Nationaux et Internationaux**

* **Service National**:
  + Durée moyenne du trajet: 157,24 minutes
  + Annulations: 92,02K
  + Retards au départ: 785,51K
  + Retards à l'arrivée: 327,47K
* **Service International**:
  + Durée moyenne du trajet: 261,81 minutes
  + Annulations: 4,64K
  + Retards au départ: 56,96K
  + Retards à l'arrivée: 18,78K

Le service national, malgré une durée de trajet moyenne plus courte, rencontre beaucoup plus de problèmes en termes de retards et d'annulations comparé au service international. Cela indique que les services nationaux nécessitent des améliorations ciblées pour atteindre une meilleure performance.

**Conclusion**

L'analyse révèle des disparités significatives entre différentes gares et services ferroviaires. Paris Montparnasse et Paris Lyon sont les gares les plus problématiques avec le plus grand nombre de retards et d'annulations, tandis que Madrid montre une performance quasi exemplaire.

Les services nationaux souffrent de plus de problèmes comparés aux services internationaux, bien que la durée moyenne des trajets soit plus courte. Des efforts ciblés sur la réduction des retards et des annulations dans les services nationaux et les gares les plus affectées sont essentiels pour améliorer la ponctualité et la satisfaction des voyageurs.

En somme, pour une amélioration globale de la performance des services ferroviaires, il est crucial de se concentrer sur les gares problématiques et les services nationaux afin de réduire les retards et les annulations, tout en maintenant les performances des services internationaux.