

## Sommaire

<b>Journal de Bord : Dashboard Immobilier aux États-Unis.....</b>	<b>2</b>
Problématique.....	2
Choix des métriques clés du dashboard.....	2
Population et salaire (relation affordability).....	3
Bases de données.....	3
Onglets.....	3
Remarques sur les bases de données.....	4
1. Loyers (Zillow).....	4
2. Revenus.....	4
BLS.....	4
IRS.....	4
Étapes du Projet.....	5
Introduction et Objectifs du Projet (Octobre 2025).....	5
Étape 1 – Choix des sources de données (Octobre 2025).....	5
Étape 2 – Nettoyage et fusion des données (Octobre 2025).....	5
Étape 3 – Mise en place de l'application streamlit (Novembre 2025).....	6
Étape 4 – Expérience utilisateur.....	6

# Journal de Bord : Dashboard Immobilier aux États-Unis

Auteur : BGHniyonkuru

Date de création : Octobre 2025

Date de mise à jour : Décembre 2025

Projet GitHub : [USA Real Estate-Dashboard](#)

Ce journal retrace l'ensemble des étapes du projet, des recherches initiales à la version finale du dashboard. Il met en lumière les choix techniques, les difficultés rencontrées, les innovations apportées et les optimisations réalisées. Le projet a évolué d'une simple visualisation de données immobilières à un outil interactif personnalisé, avec une prévision dynamique à 5-10 ans, intégrant des calculs réalistes d'abordabilité (PITI + HOA) et une granularité fine au niveau ZIP Code enrichie par des noms de villes lisibles.

## Problématique

Un ami qui vit aux États-Unis m'a exposé la difficulté de trouver une ville où il pourrait investir en achetant ou en louant une maison. Je lui ai donc proposé de modéliser une application qui lui permette d'avoir une vue globale de ce qui se fait en termes de logement aux États-Unis, et aussi de pouvoir avoir une prédiction sur 5 ans ou 10 ans afin qu'il vérifie que son investissement va être réussi ou pas.

## Choix des métriques clés du dashboard

Un bon dashboard devrait traquer des indicateurs essentiels pour évaluer la santé du marché immobilier. Voici une sélection prioritaire, inspirée des pratiques standard :

- **Prix des logements** : Valeur médiane des maisons (ZHVI chez Zillow), indice des prix (HPI chez FHFA), prix de vente médian, prix au m².
- **Inventaire et offre** : Nombre de listings actifs, nouvelles listings, listings en attente (pending).
- **Ventes et demande** : Volume des ventes, ratio vente/prix listé, pourcentage de ventes sous/au-dessus du prix demandé.
- **Durée sur le marché** : Jours pour passer en pending ou clôturer une vente (days on market).
- **Loyers et accessibilité** : Indice des loyers (ZORI), paiements hypothécaires estimés, revenu nécessaire pour l'accessibilité (part de revenu allouée au logement).

- **Autres** : Indice de la chaleur du marché (market heat index), prévisions de prix, taux de réduction de prix, indicateurs d'investissement comme le NOI (Net Operating Income) ou le cap rate pour les investisseurs.

**Astuces** : Utilisez des graphiques comme des lignes temporelles pour les tendances, des cartes pour la géographie, et des tableaux pour comparer les États/métropoles.

## Population et salaire (relation affordability)

Pour étudier la population qui peut se permettre d'acheter, louer ou vendre :

- **Revenu médian des ménages (Median Household Income)** : Revenu annuel médian par ménage, idéal pour les calculs d'accessibilité.
- **Revenu par habitant (Per Capita Income)** : Pour une vue individuelle, utile si vous analysez les ménages solos.
- **Salaire moyen/médian par occupation** : Par secteur (ex. : tech, santé) pour segmenter la population active.
- **Croissance des salaires** : Variation annuelle pour prédire les tendances d'accessibilité.
- **Métriques dérivées d'accessibilité** :
  - Revenu nécessaire pour l'achat/location (income needed to afford).
  - Part du revenu consacré au logement (affordability ratio).
  - Années pour économiser un apport (years to save for down payment).

Ces données sont souvent annuelles ou trimestrielles, avec une granularité allant du national au ZIP Code. Pour l'intégration, fusionnez-les avec vos datasets immobiliers via des clés communes (ex. : code ZIP, nom de métro).

## Bases de données

Les bases de données privilégiées sont gratuites et fiables. Parmi celles utilisées :

1. **Census** : Données démographiques et revenus.
2. **Zillow Research** : Fichiers en CSV actualisés tous les 16 du mois sur le site Zillow.
3. **RedFin News** : Fichiers TSV.gz hebdomadaires (mercredis) et mensuels sur le site RedFin.
4. **FHFA House Price** : Données sur les indices de prix des maisons sur le site FHFA.
5. **Realtor.com Research** : Données sur les listings et le marché.
6. **Autres** : BLS (salaires), IRS (revenus fiscaux).

## Onglets

- **Page d'accueil** : Explicatif de ce qu'on peut faire avec le dashboard.
- **Dashboard** :
  1. Vue nationale
  2. Vue personnelle

- **Prédiction**

## Remarques sur les bases de données

### 1. Loyers (Zillow)

- **Home values** : The total monthly payment includes the mortgage payment, homeowner's insurance, property taxes, and maintenance costs worth 0.5% of the home's value. The home value is estimated using smoothed and seasonally adjusted ZHVI.
- **Home values forecasts** : Zillow Home Value Forecast (ZHVF): A month-ahead, quarter-ahead and year-ahead forecast of the Zillow Home Value Index (ZHVI). ZHVF is created using the all homes, mid-tier cut of ZHVI and is available both raw and smoothed, seasonally adjusted.
- **Rentals** : Données sur les loyers.
- **For sales listings** : Données sur les listings à vendre.
- **Days on market** : Temps sur le marché.
- **Market Heat** : Indice de chaleur du marché.

### 2. Revenus

#### BLS

Data from BLS. See if the trusted data from IRS can be used instead of the BLS.

#### "all\_data\_M\_2024.xlsx"

Ce dataset couvre environ 830 occupations (métiers) par zone géographique (principalement métros/MSA, États, et national). Il est idéal pour segmenter la population active (ex. : salaires des ingénieurs ou enseignants dans une métro comme New York, pour voir qui peut s'offrir un logement). Il y a 32 colonnes, mais concentrez-vous sur ces 10-15 les plus utiles pour votre analyse.

#### IRS

#### Différence entre les deux fichiers CSV IRS sur votre image.

Ces fichiers sont des datasets IRS pour les revenus fiscaux 2022 par ZIP Code (tous États). La différence est dans le contenu des colonnes sensibles pour la confidentialité :

- **2022 (all States, includes AGI)** : Inclut l'Adjusted Gross Income (AGI) total et par catégorie (ex. : AGI pour les revenus <25k, 25-50k, etc.). AGI est le revenu brut ajusté (salaires + investissements - déductions). Choisissez celui-ci si vous voulez des détails fins sur les distributions de revenus (ex. : pour étudier les hauts revenus qui achètent des maisons chères).

- **2022 (all States, does not include AGI)** : Exclut l'AGI pour éviter les données trop précises sur les individus riches (pour protéger la vie privée dans les petits ZIP). Il inclut toujours le nombre de déclarations, exemptions, etc., mais sans les montants d'AGI. Choisissez celui-ci si vous avez besoin seulement de counts (ex. : nombre de ménages par tranche de revenu), ou pour des analyses plus légères.

**Recommandation** : Téléchargez le premier ("includes AGI") pour plus de valeur dans votre dashboard – il permet de calculer des médians/moyennes personnalisés. Les deux sont en CSV, ~30-50 Mo, et couvrent tous les ZIP US.

## Étapes du Projet

### Introduction et Objectifs du Projet (Octobre 2025)

Le projet vise à créer un dashboard interactif analysant l'accessibilité immobilière aux États-Unis, en croisant données sur les prix des logements, les loyers, les salaires et les revenus réels des ménages. L'objectif est de répondre à la question : "Où une personne avec un salaire donné peut-elle s'offrir un logement aujourd'hui et dans 5 ans ?".

**Innovation clé** : Premier dashboard étudiant intégrant une prévision personnalisée basée sur des paramètres dynamiques (salaire, apport, horizon, inflation), avec calcul réaliste du coût total d'achat (principal, intérêts, taxes, assurance, HOA).

#### Métriques clés du dashboard (calculées dynamiquement) :

- Income\_Needed\_Buy : Calculé via PITI + HOA.
- Income\_Needed\_Rent =  $(ZORI \times 12) / 0.3$ .
- Affordability\_% =  $(Avg\_AGI / Income\_Needed\_X) \times 100$ .
- Location = Ville + " (" + Metro + ")" (lisibilité maximale).

### Étape 1 – Choix des sources de données (Octobre 2025)

Tous les jeux de données mentionnées avant n'ont pas toutes été exploitées car les fichiers étaient lourds.

- Zillow
- BLS
- IRS

### Étape 2 – Nettoyage et fusion des données (Octobre 2025)

- Fusion de 3 datasets sur la clé ZIP (41 000+ lignes)
- Création des colonnes clés :
  - Income Needed Buy → calculé dynamiquement
  - Income\_Needed\_Rent =  $(ZORI \times 12) / 0.3$
  - Affordability\_% =  $(Avg\_AGI / Income\_Needed\_X) \times 100$

- Location = Ville + " (" + Metro + ")" → lisibilité maximale

### Étape 3 – Mise en place de l'application streamlit (Novembre 2025)

- Codage des onglets
- Connection avec les jeux des données

### Étape 4 – Expérience utilisateur

- **Explications claires** sur l'affordability (texte + colonnes pédagogiques)
- **Hover informatif** sur tous les graphiques
- **Message d'absence de résultat** avec conseil (ex: "Essayez de réduire l'apport")
- **Page d'accueil** expliquant l'objectif + l'innovation (prévision personnalisée)