

---

# **PROJET FIN D'ANNÉE**

---

**Présenté par :**

**Abdelkader Arras**

**Yassine El Maghraoui**

**Anouar Chalabi**

**Membres du jury :**

**Mr Kamal Najem**

**Mme Asmae Ouhmida**

# INTRODUCTION

# CONTENU

- 1 **Introduction**
- 2 **Contexte général du projet**
- 3 **Problématique**
- 4 **Objectifs**
- 5 **Analyse exploratoire et prétraitement des données**
- 6 **Modélisation**
- 7 **Réalisation d'application décisionnelle**
- 8 **Conclusion et perspectives**

# CONTEXTE GÉNÉRAL DU PROJET

Le secteur du e-commerce connaît une croissance rapide, accompagnée d'une concurrence de plus en plus intense entre les plateformes. Dans ce contexte, fidéliser les clients devient un véritable défi, car ces derniers disposent de nombreuses alternatives. Le churn client a un impact direct sur le chiffre d'affaires et la rentabilité des entreprises, rendant son anticipation essentielle. L'exploitation des données clients, combinée aux techniques de machine learning, constitue aujourd'hui un levier clé pour mieux comprendre les comportements et mettre en place des stratégies de rétention efficaces.

# PROBLEMATIQUE

**Comment prédire efficacement le churn des clients e-commerce tout en fournissant des explications interprétables et exploitables, afin d'aider les décideurs à mettre en place des stratégies de rétention ciblées et personnalisées ?**

# OBJECTIFS

- Identifier les clients à risque de churn
- Comprendre les facteurs expliquant le churn
- Proposer des actions de rétention ciblées
- Développer une application interactive d'aide à la décision

# PRÉSENTATION DES DONNÉES

# PRÉSENTATION DES DONNÉES

Catégorie	Variable	Description	Type de Donnée
Identifiant	CustomerID	Identifiant unique du client.	Qualitatif (ID)
Cible	Churn	Indicateur de départ du client.	Qualitatif (Binaire)
Démographie	Tenure	Ancienneté du client au sein de l'organisation.	Quantitatif Discret
Démographie	CityTier	Niveau de la ville du client (Tier 1, 2, 3).	Qualitatif Ordinal
Démographie	Gender	Genre du client.	Qualitatif Nominal
Démographie	MaritalStatus	Statut marital du client.	Qualitatif Nominal
Démographie	NumberOfAddress	Nombre total d'adresses ajoutées par le client.	Quantitatif Discret
Comportement	PreferredLoginDevice	Appareil de connexion préféré.	Qualitatif Nominal
Comportement	HourSpendOnApp	Nombre d'heures passées sur l'application.	Quantitatif Continu
Comportement	NumberOfDevices	Nombre total d'appareils enregistrés.	Quantitatif Discret
Comportement	PreferredOrderCat	Catégorie de commande préférée le mois dernier.	Qualitatif Nominal
Comportement	SatisfactionScore	Score de satisfaction du client.	Quantitatif Discret

Catégorie	Variable	Description	Type de Donnée
Comportement	Complain	Indicateur si une plainte a été soulevée.	Qualitatif Binaire
Transactionnel	WarehouseToHome	Distance entrepôt-domicile (km).	Quantitatif Continu
Transactionnel	PreferredPaymentMode	Mode de paiement préféré.	Qualitatif Nominal
Transactionnel	OrderAmountHikeFromLastYear	Pourcentage d'augmentation des commandes.	Quantitatif Continu
Transactionnel	CouponUsed	Nombre de coupons utilisés le mois dernier.	Quantitatif Discret
Transactionnel	OrderCount	Nombre de commandes passées le mois dernier.	Quantitatif Discret
Transactionnel	DaySinceLastOrder	Jours écoulés depuis la dernière commande.	Quantitatif Discret
Transactionnel	CashbackAmount	Montant moyen de cashback le mois dernier.	Quantitatif Continu

# **ANALYSE EXPLORATOIRE DES DONNÉES ET PRÉTRAITEMENT**

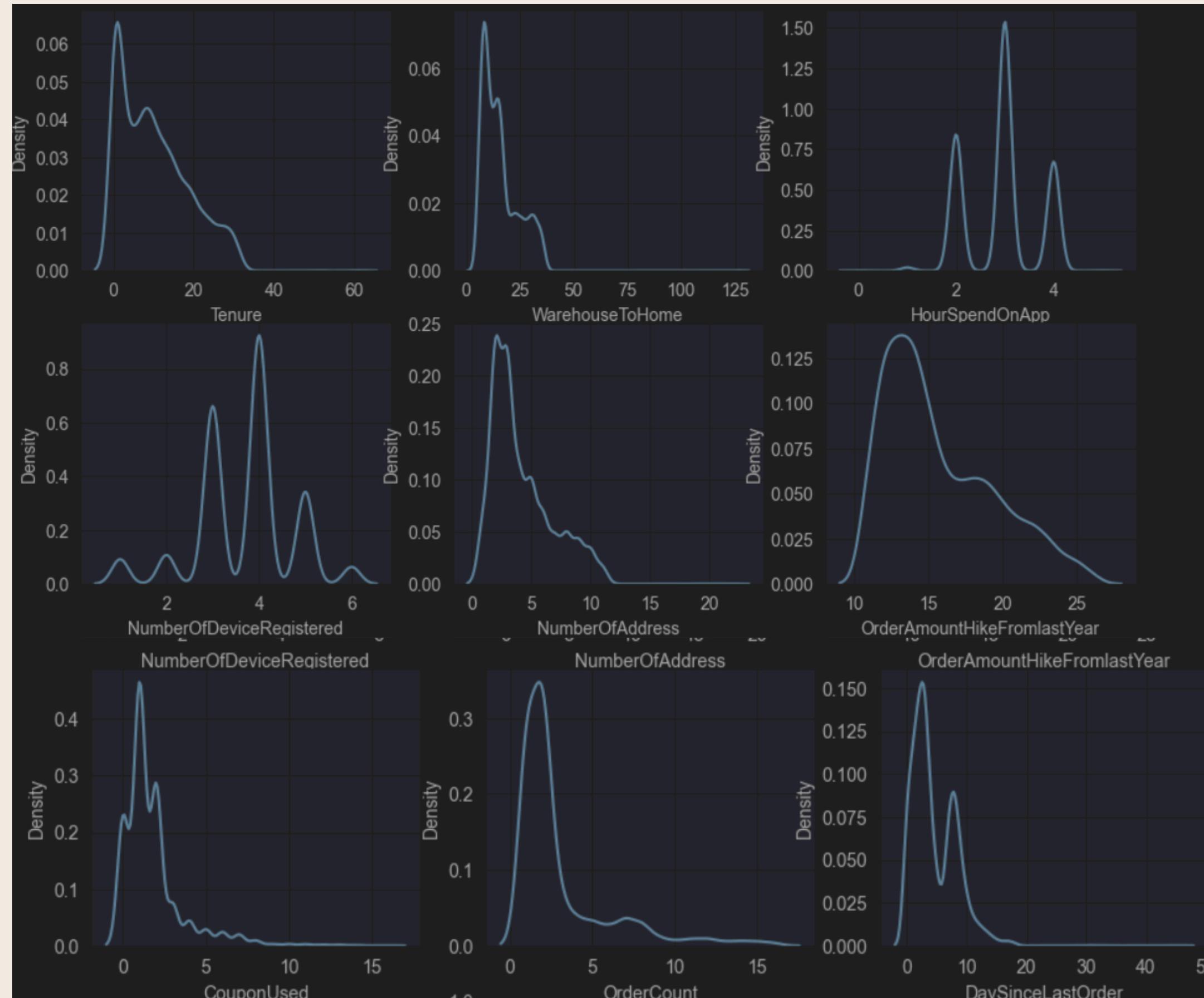
# VALEURS MANQUANTES

Tenure	264
WarehouseToHome	251
HourSpendOnApp	255
OrderAmountHikeFromLastYear	265
CouponUsed	256
OrderCount	258
DaySinceLastOrder	307
<b>dtype:</b>	<b>int64</b>

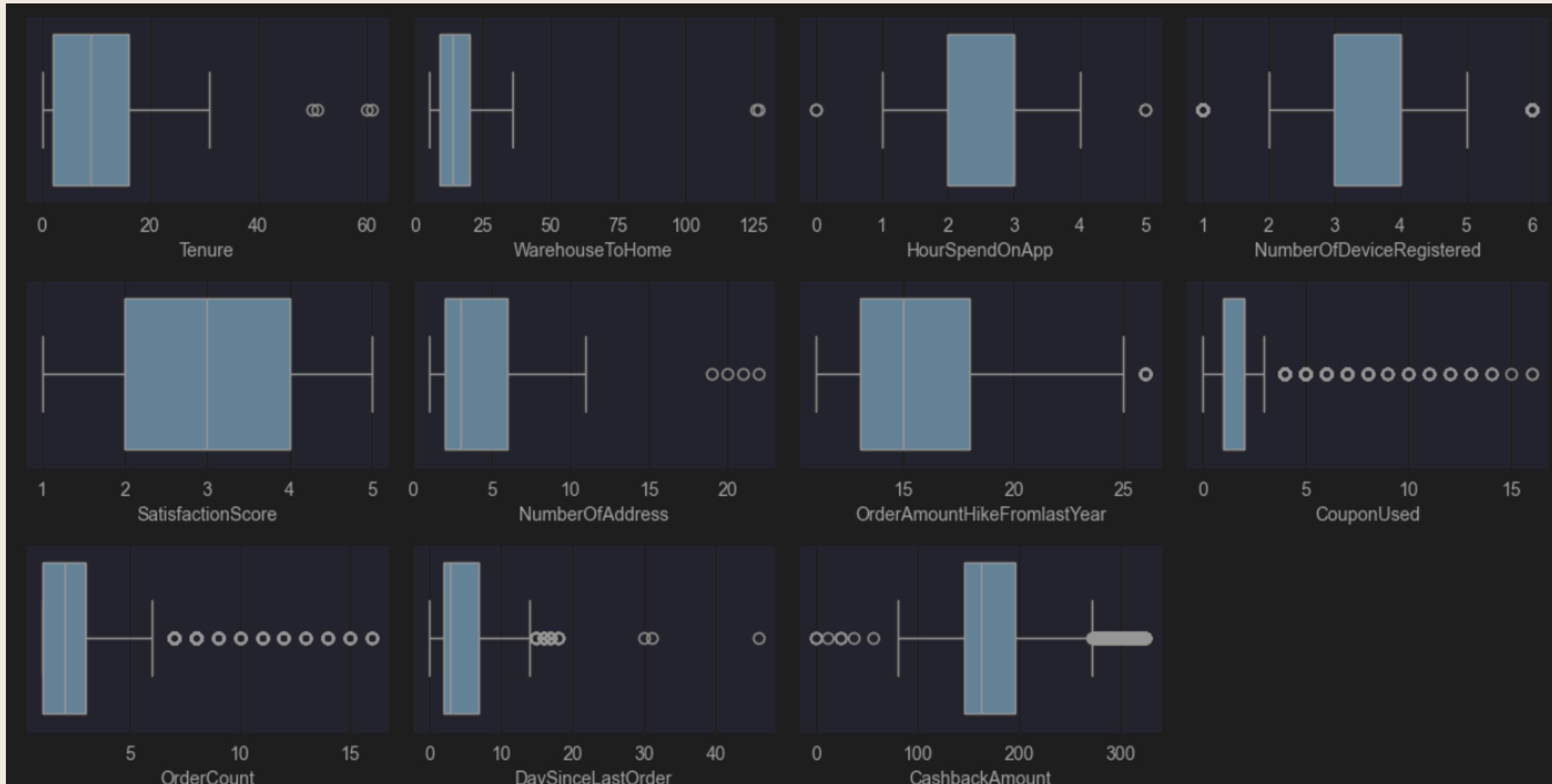
Tenure	4.69
WarehouseToHome	4.46
HourSpendOnApp	4.53
OrderAmountHikeFromLastYear	4.71
CouponUsed	4.55
OrderCount	4.58
DaySinceLastOrder	5.45

- **1 856 valeurs manquantes détectées dans le jeu de données**
- **1 856 lignes concernées, soit  $\approx 33\%$  des observations**
- **Les variables les plus affectées sont liées au comportement d'achat et à l'engagement client**

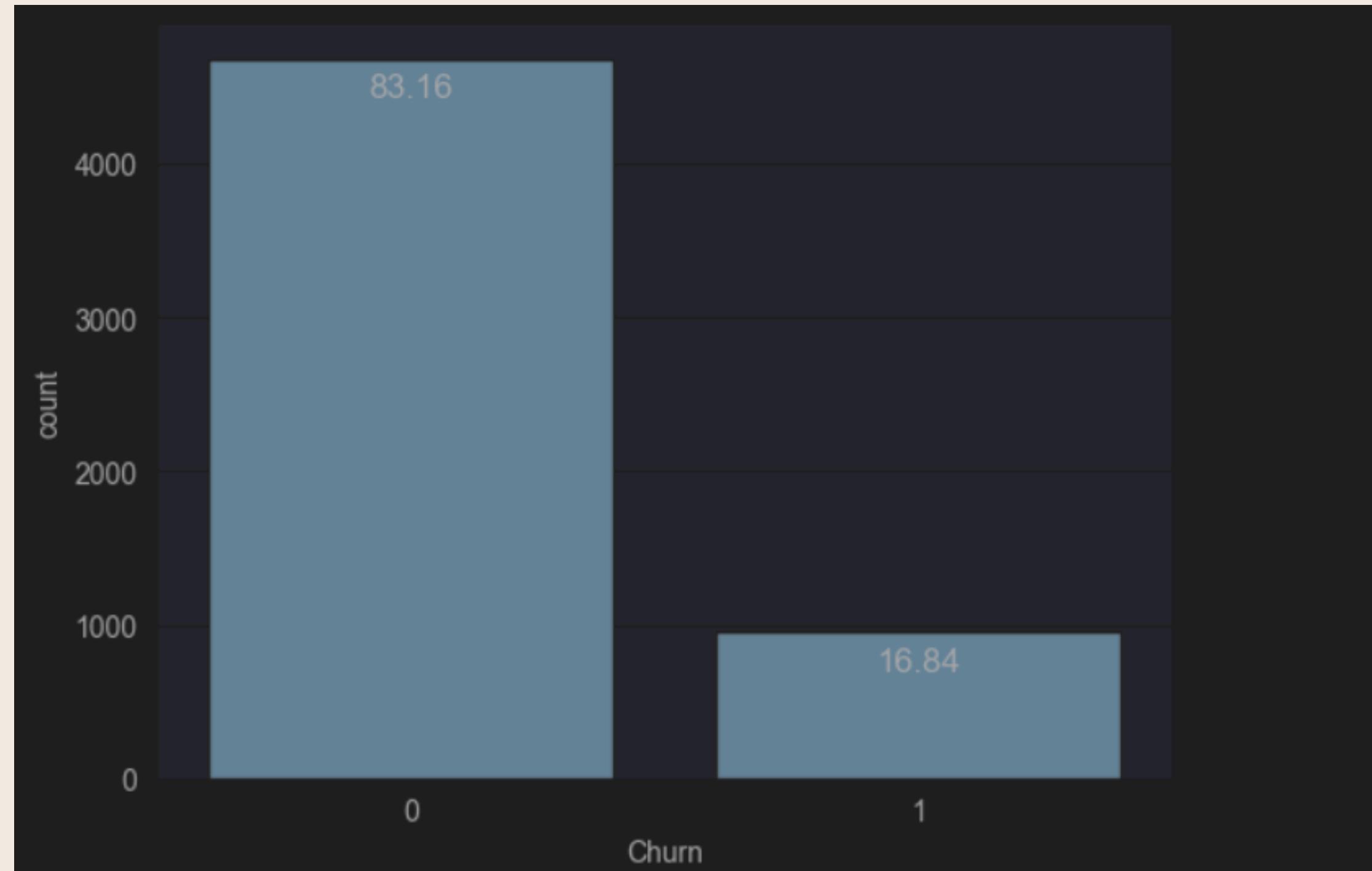
# DISTRIBUTION DES VARIABLES



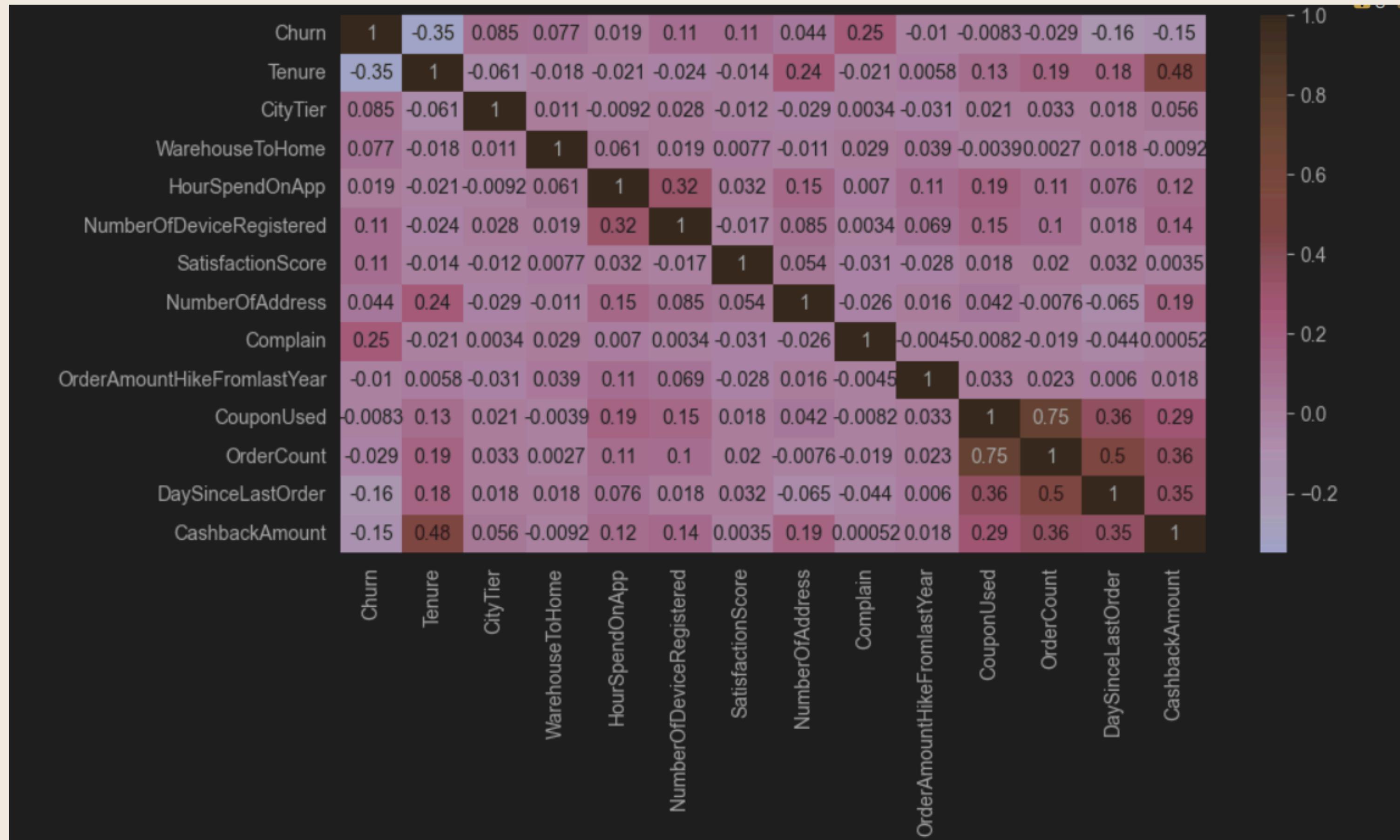
# VALEURS ABERRANTES(OUTLIERS)



# ANALYSE DE LA VARIABLE CIBLE (CHURN VS NON-CHURN)



# ANALYSE DES CORRÉLATIONS ENTRE VARIABLES



# MODÉLISATION

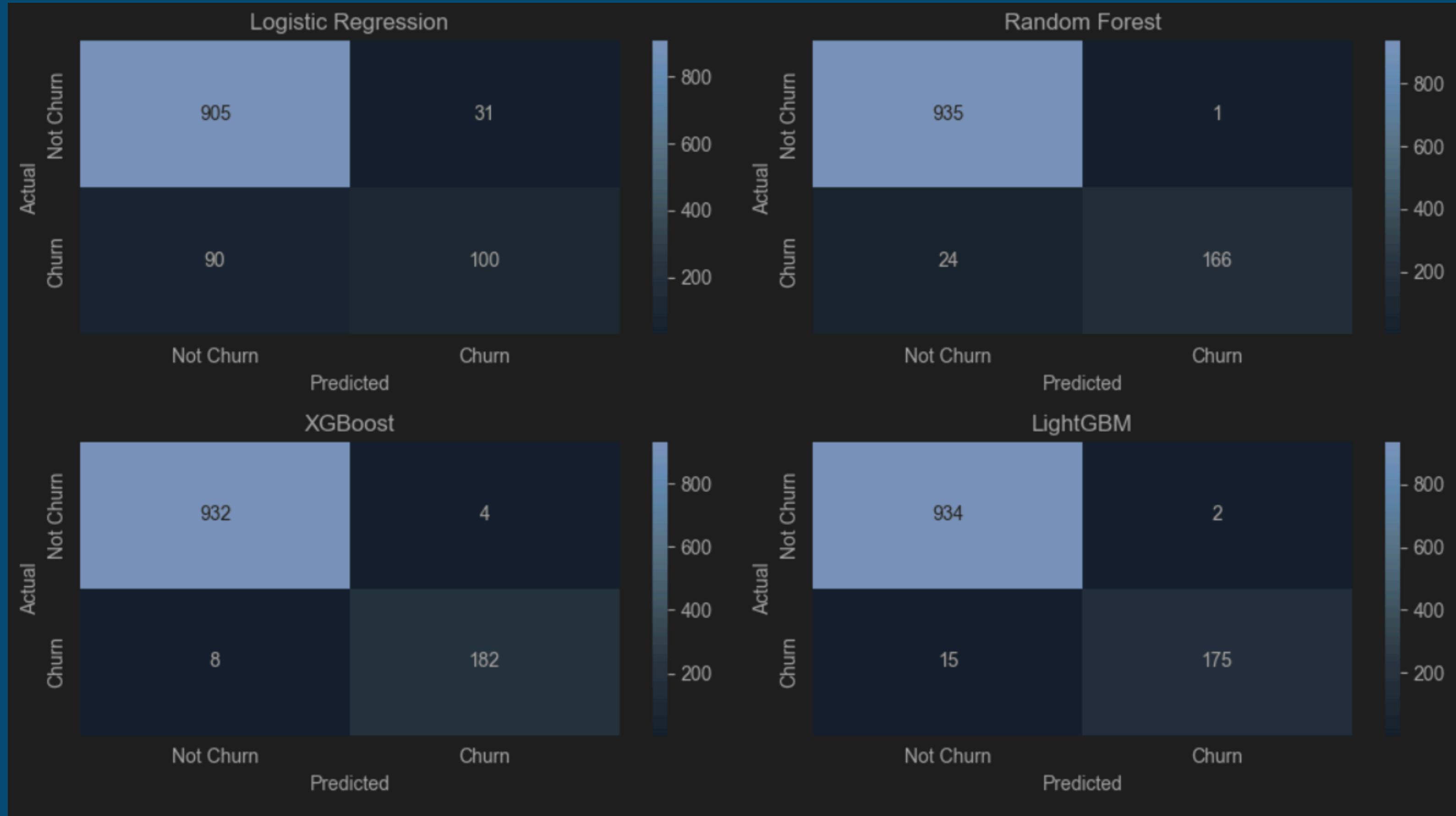
# CHOIX DES MODELES

- **Regression logistique**
- **RaNDom Forrest**
- **XGBoost**
- **LIGHTGBM**

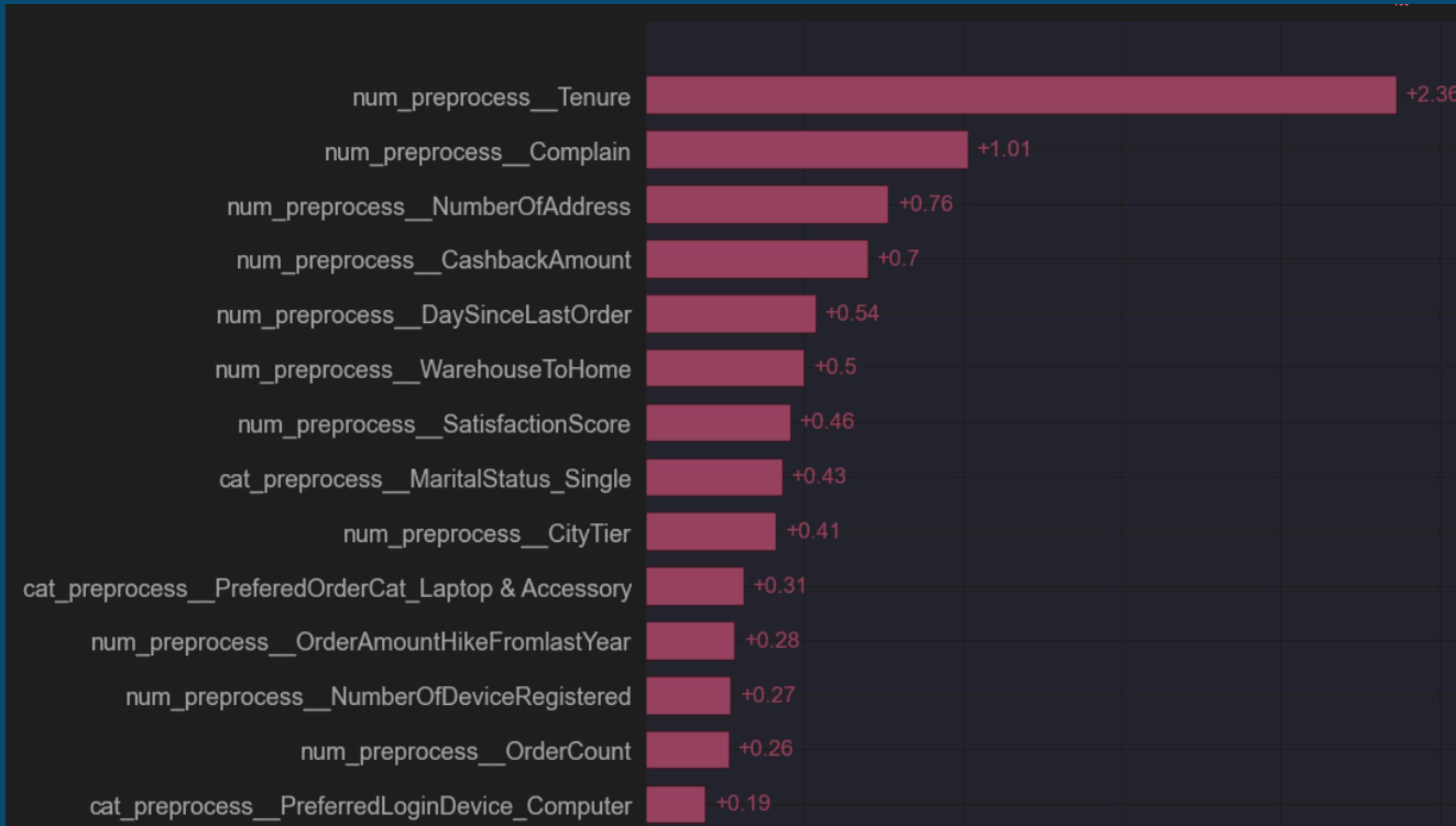
# RÉSULTATS ET COMPARAISON DES MODÈLES

#	Model	CV Average F1	Train Acc	Test Acc	Train F1	Test F1
0	Logistic Regression	0.627823	0.898757	0.892540	0.645412	0.623053
1	Random Forest	0.811769	1.000000	0.977798	1.000000	0.929972
2	XGBoost	0.866715	1.000000	0.989343	1.000000	0.968085
3	LightGBM	0.843233	0.998446	0.984902	0.995367	0.953678

# RÉSULTATS ET COMPARAISON DES MODÈLES



# ANALYSE GLOBALE DES VARIABLES INFLUENTES AVEC SHAP



# RÉALISATION D'APPLICATION DÉCISIONNELLE

# CONCLUSION ET PERSPECTIVES

**MERCI POUR VOTRE ATTENTION**