LendSense : A Smart Data-Driven Credit Risk Model

Prepared By :
FEBRIYAN CHANDRA

Final Task - Data Science Project Based Internship at ID/X Partners

Table of Content

Business Understanding

Company Profile

Project Overview

Problem Breakdown

Company Profile

Founded: 2005
Industry: Data & Analytics Consulting
Services: Data Science, Business Intelligence, Big Data, AI Solutions
Mission: Helping businesses optimize data for better decision-making
Website: www.idxpartners.com

Project Overview

This project involves cross-departmental collaboration to develop a technology-driven solution for the company. The goal is to build **a predictive model for credit risk assessment using the given dataset, which contains data from both approved and rejected loan applications.**

Problem Breakdown

Current Condition
Lending companies face **high default rates** due to **traditional risk assessment methods.**

Ideal Condition
An accurate credit risk prediction model enables better loan decisions, reducing defaults and improving profitability.

Gap / Problem
Data-driven solutions are not fully utilized, leading to less accurate credit risk evaluations.

Credit Risk Modelling Workflow

- Data Exploration
- Data Preparation & Preprocessing
- Modelling
- Weight of Evidence
- New Features

Data Exploration

Objective: Understanding the structure and distribution of data while identifying initial trends that may affect credit risk before further analysis.

Workflow

1. Check number of columns and data types of features
2. Create a new column based on the Loan Status column that will be target variable
3. Check shape of data
4. Show columns that have more than 70% missing values

Insight

Data Structure :
Total Observations:
466,285 rows, 75 features

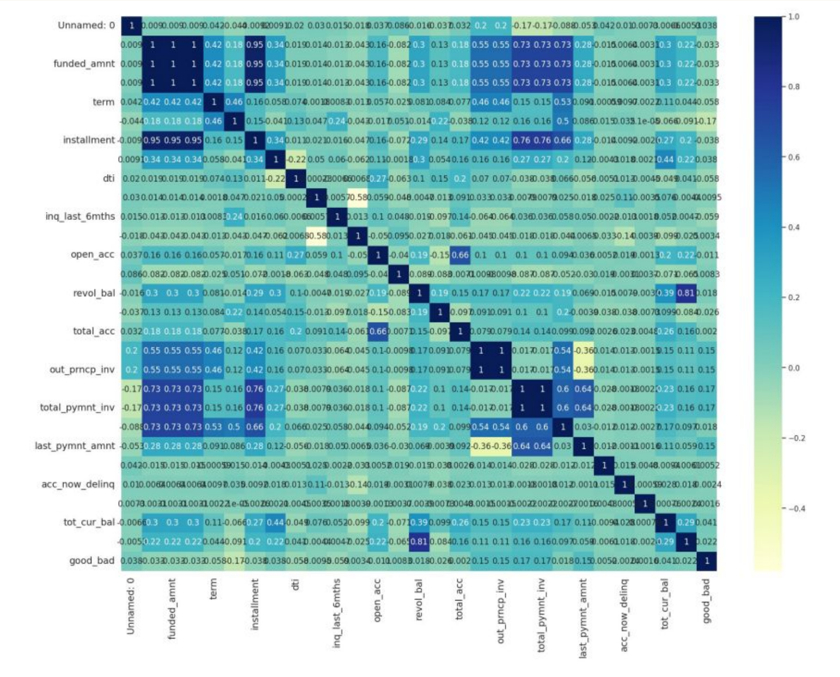
There are **20 columns** with more than **70% missing values.**

Data Preparation & Preprocessing

Dropping of irrelevant columns and rows with missing values (33 columns)

Reason for Removal	Columns
Redundant or Identifiers	id, member_id, sub_grade
Unstructured or Irrelevant Data	emp_title, url, desc, title
Privacy Concerns	zip_code
Future or Post-loan Data (Not Useful for Predictions)	next_pymnt_d, recoveries, collection_recovery_fee, total_rec_prncp, total_rec_late_fee
Missing or Incomplete Data	mths_since_last_record, mths_since_last_major_derog, mths_since_rcnt_il
Joint Loan Data (Not in Scope)	annual_inc_joint, dti_joint, verification_status_joint
Low Impact or Less Relevant Features	open_acc_6m, open_il_6m, open_il_12m, open_il_24m, total_bal_il, il_util, open_rv_12m, open_rv_24m, max_bal_bc, all_util, inq_fi, total_cu_tl, inq_last_12m
Constant or Low Variability Features	policy_code

Data Preparation & Preprocessing



Key Findings

<u>High Correlations:</u>

- **Funded Amount, Installment, and Total Payment (≥ 0.73)** → Higher loans lead to higher installments and total payments.
- **Total Accounts & Open Accounts (~0.66)** → More open accounts mean a higher total number of accounts.
  Business Impact: Key indicators for credit behavior and risk evaluation.

<u>Weak Correlations:</u>

- **Debt-to-Income Ratio (DTI) has low correlation with financial variables.**
  Business Impact: DTI alone isn't a strong credit risk determinant.

<u>Target Variable (good_bad) Relationship:</u>

- **Total Payment (~0.15) with good_bad** → Higher payments lower default risk.
  Business Impact: Useful predictor for creditworthiness.

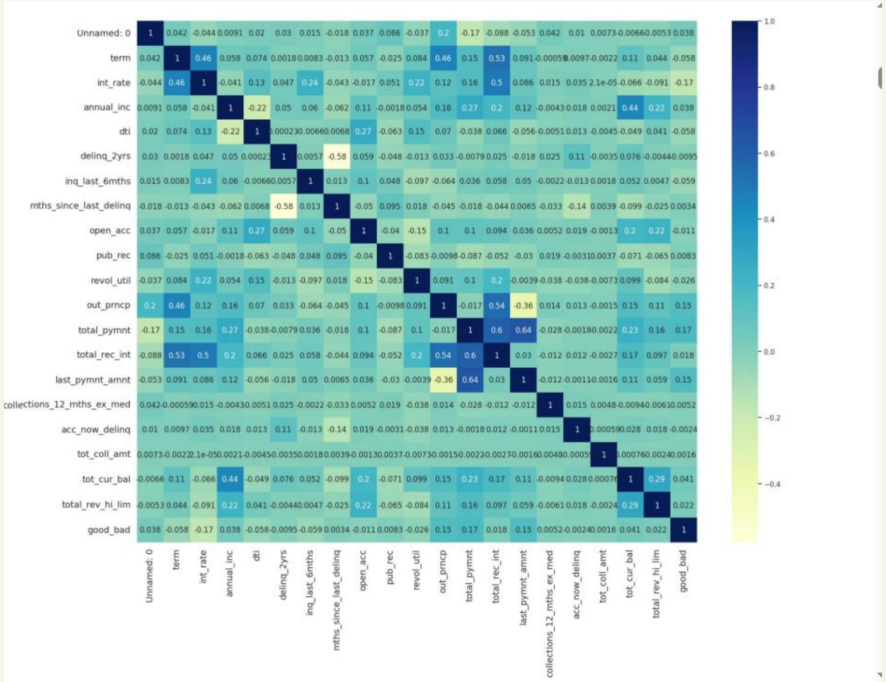
Data Preparation & Preprocessing

Dropping multicollinear features (6 columns)

Columns
Loan Amount
Revolving Balance
Funded Amount
Investor Funded Amount
Installment
Total Payment

Data Preparation & Preprocessing

After Dropping multicollinear features (6 columns)



Key Findings

High Correlations:

- **Total Payment & Last Payment Amount (~0.64)** → Higher total payments are strongly related to larger last payments.
- **Term & Outstanding Principal (~0.46)** → Longer loan terms tend to have higher remaining principal balances.
- **Total Current Balance & Total Revolving High Limit (~0.49)** → Higher credit limits are associated with higher current balances.
- Business Impact: These variables are key indicators of loan repayment behavior and credit risk assessment.

Weak Correlations:

- **Public Records, Collections, and Accounts Delinquency** have very low correlations with other financial variables. Business Impact:These factors alone may not significantly influence creditworthiness and require further analysis.
- **Target Variable (`good_bad`) Relationship: Total Payment (~0.17) & Good/Bad Status** → Higher payments are associated with lower credit risk.
- **Revolving Utilization (~-0.26) & Good/Bad Status** → Higher credit utilization correlates with a higher risk of being classified as a bad debtor.
- **Months Since Last Delinquency (~-0.17) & Good/Bad Status**→ Longer time since last delinquency correlates with lower credit risk. Business Impact:Payment behavior and credit utilization are crucial in predicting loan defaults.

Final Dataframe

Column Name	Description
Index	Unique index number for each data entry
Loan Term	Loan duration (in months)
Interest Rate	Loan interest rate (%)
Credit Grade	Borrower's credit risk classification
Employment Length	Borrower's employment length (in years)
Home Ownership Status	Borrower's home ownership status
Annual Income	Borrower's annual income
Verification Status	Income verification status by the lender
Payment Plan	Whether the borrower has an active payment plan
Loan Purpose	The primary reason for taking the loan
State Address	Borrower's state or city of residence
Debt-to-Income Ratio (DTI)	Ratio of debt to income
Delinquencies (Last 2 Years)	Number of late payments in the last 2 years
Inquiries (Last 6 Months)	Number of credit inquiries in the last 6 months

Column Name	Description
Months Since Last Delinquency	Months since the last delinquency
Open Credit Accounts	Number of active credit accounts
Public Records	Number of negative records in the credit report
Revolving Credit Utilization	Percentage of revolving credit being used
Initial Listing Status	Initial loan listing status (e.g., offered to investors)
Outstanding Principal Balance	Remaining principal balance on the loan
Total Payment Made	Total amount paid by the borrower
Total Interest Received	Total interest earned by the lender
Last Payment Amount	Last payment amount made by the borrower
Collections (Last 12 Months)	Number of accounts sent to collections in the last year
Application Type	Type of loan application (individual or joint)
Accounts Currently Delinquent	Number of accounts currently overdue
Total Collection Amount	Total amount in collections
Total Current Balance	Total current balance across all accounts
Total Revolving Credit Limit	Total revolving credit limit of the borrower
Loan Status (Good/Bad)	Loan status (good or bad) - target variable
Months Since Loan Issued	Months since the loan was issued
Months Since Last Payment	Months since the last payment was made
Months Since Last Credit Pull	Months since the last credit check
Months Since Earliest Credit Line	Months since the borrower's first credit account

Weight of Evidence

What is WOE?

WOE is a technique used in credit risk modeling to **transform variables into a log-scale representation based on the likelihood of an event (e.g., default vs. non-default)**. It helps identify patterns in data, improve predictive modeling, and enhance decision-making in risk management.

Monotonic Trends (Strong Predictors):

- Variables with **steadily increasing or decreasing WOE** indicate a strong relationship with credit risk.

Business Insight:

- **Higher income → Lower default risk**, supporting income-based risk assessments.
- **Higher loan amounts → Higher default risk**, indicating the need for stricter approval criteria for large loans.

Non-Monotonic Patterns (Needs Refinement):

- Some variables show **fluctuating WOE**, which may indicate mixed risk behavior or poor binning.

Business Insight:

- **Loan tenure shows inconsistencies**, suggesting that mid-range tenures might have hidden risks.
- Adjusting loan policies for certain tenure ranges could **optimize default rates**.

Unstable WOE (Potential Data Issues):

- Sharp WOE variations between bins suggest **low sample sizes or noisy data**, requiring re-binning or removal.

Business Insight:

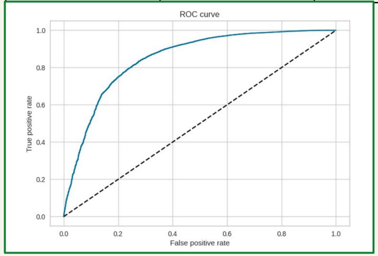
- **Employment length has irregular jumps**, meaning default risk isn't clearly defined for short vs. long-term employees.
- **HR-based segmentation** or additional financial indicators may be needed for better risk assessment.

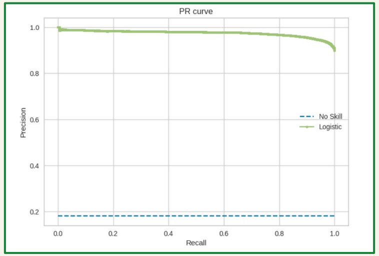
New Features

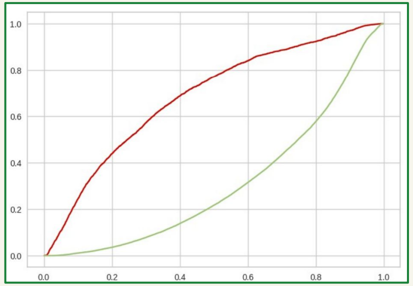
Feature Name	Description
Home Ownership	Borrower's home ownership status, such as "Rent," "Own," or "Mortgage."
Verification Status	Status of borrower's information verification, such as "Verified" or "Not Verified."
Purpose of Loan	The purpose of the loan, e.g., "Debt Consolidation," "Home Improvement," etc.
State	The borrower's state of residence, represented as a two-letter code (e.g., CA, NY).
Loan Term	The duration of the loan in months, such as "36 months" or "60 months."
Total Interest Received	The total amount of interest paid by the borrower over the loan period.
Total Revolving Credit Limit	The total maximum revolving credit limit of the borrower.
Total Payment	The total amount of payments made by the borrower so far.
Interest Rate	The loan interest rate as a percentage.
Debt to Income Ratio	The ratio of the borrower's total debt to their annual income.
Annual Income	The borrower's annual income.
Credit Inquiries Last 6 Months	The number of times the borrower applied for credit in the past six months.
Total Current Balance	The total current balance across all of the borrower's credit accounts.
Months Since Last Credit Pull	The number of months since the borrower's last credit report was pulled.
Months Since Loan Issued	The number of months since the loan was issued.

Modelling

Class	Precision	Recall	F1-Score	Support
0	0.28	0.77	0.41	3,640
1	0.97	0.78	0.87	32,698
Accuracy	-	-	0.78	36,338
Macro Avg	0.63	0.78	0.64	36,338
Weighted Avg	0.9	0.78	0.82	36,338







Key Metrics:

KS Score: 0.558 → Strong Differentiation (Above industry benchmark of 0.4 - 0.5)

Precision (Bad Loans - Class 0): 0.28 → Higher false positives may lead to missed lending opportunities.

Recall (Bad Loans - Class 0): 0.77 → Effective in detecting high-risk borrowers.

Precision (Good Loans - Class 1): 0.97 → Ensures high confidence in approving low-risk customers.

F1-Score (Macro Avg): 0.64 → Model favors good loans, suggesting a need for rebalancing.

Business Insights:

Risk Mitigation: High KS Score indicates strong ability to differentiate credit risk, reducing default rates.

Revenue Optimization: Low precision on bad loans means potential lost opportunities—some rejected borrowers may actually be creditworthy.

Operational Efficiency: Reliable classification of good loans ensures faster approvals, improving customer experience.

Strategic Action: Adjust credit policies or refine the model to reduce false positives and improve risk-adjusted returns.

Key Insights & Strategic Takeaways

Business Insights

Conclusion

Business Insights

Risk-Based Pricing & Credit Policy Optimization

The model's high KS Score (0.558) confirms strong risk differentiation, allowing financial institutions to implement tiered interest rates based on borrower risk levels.

Strategic Action: Lower interest rates for high-creditworthy customers to improve customer acquisition while adjusting risk premiums for potentially risky borrowers..

Improving Loan Approval Efficiency

With high precision for good loans (0.97), institutions can confidently approve creditworthy customers, leading to faster disbursements and better customer experience.

Strategic Action: Automate approvals for low-risk segments, reducing manual review time and operational costs.

Enhancing Default Prevention Strategies

The model has 77% recall for bad loans, effectively detecting high-risk borrowers but with false positives (low precision for bad loans: 0.28).

Strategic Action: Introduce behavioral scoring and alternative data (transaction history, mobile data, etc.) to refine high-risk borrower segmentation.

Balancing Business Growth & Risk Control

Expand lending with tailored loan products for different risk segments (secured vs. unsecured).

Strategic Action: Develop dynamic lending policies → Adjust risk thresholds based on macroeconomic trends and market conditions.

Conclusion

Business Understanding & Problem Framing	Data Processing & Feature Engineering	Credit Risk Modelling & Evaluation	Strategic Business Recommendations	Future Enhancements
Challenge: Traditional credit risk assessment leads to high default rates. **Goal:** Develop a predictive model to optimize loan approvals, reduce defaults, and drive profitability.	**Key Insights:** ✔ Loan amount, installments, & credit behavior → Highly correlated. ✔ DTI ratio → Weak predictor of credit risk. ✔ High total payments & low credit utilization → Strong indicators of lower default risk. **Action:** Removed irrelevant/multicollinear features & optimized variable selection.	**KS Score: 0.558** → Strong risk differentiation (above industry standard). **F1-Score (0.64)** → Model slightly favors good loans, requiring better balance. **Precision (Bad Loans: 0.28)** → High false positives, indicating missed lending opportunities. **Precision (Good Loans: 0.97)** → Ensures high confidence in approving low-risk customers.	**Risk-Based Pricing** → Adjust loan interest rates based on borrower risk levels. **Alternative Data Integration** → Use spending behavior, mobile data, and financial history for better borrower assessment. **Portfolio Expansion with Risk Control** → Develop tiered lending strategies for different risk segments. **AI-Powered Credit Decisioning** → Automate loan approvals for faster processing and reduced defaults.	• **Reduce false positives** to capture more creditworthy borrowers. • **Incorporate adaptive machine learning** to improve accuracy over time. • **Implement real-time risk monitoring** to dynamically adjust lending policies.

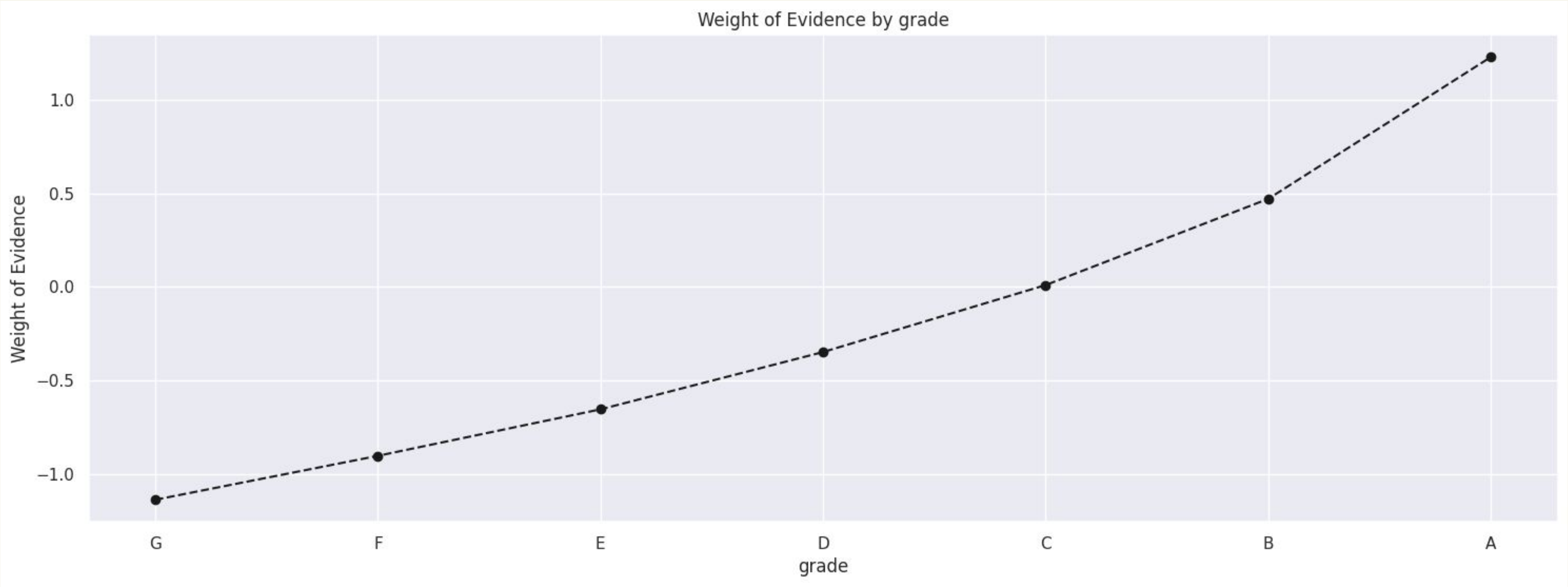
Rakamin id/x partners

THANK YOU

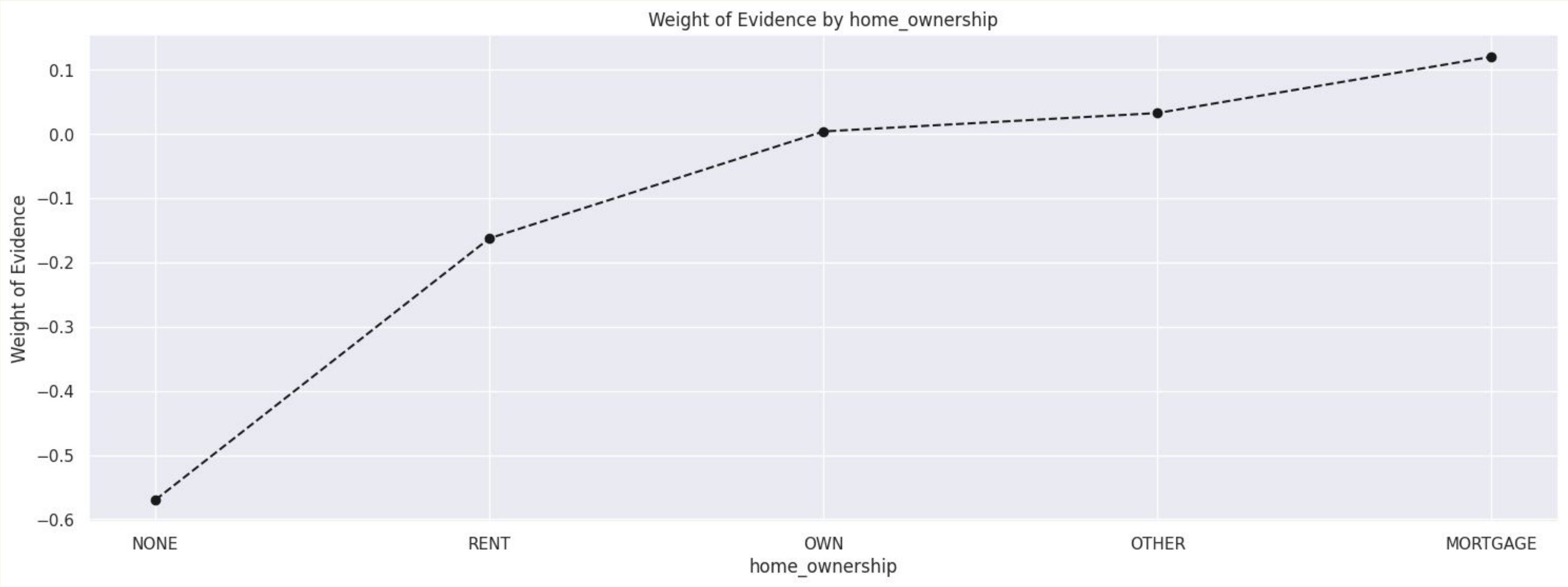
Contact Person:
FEBRIYAN CHANDRA
chanfebriyanch@gmail.com

Appendix

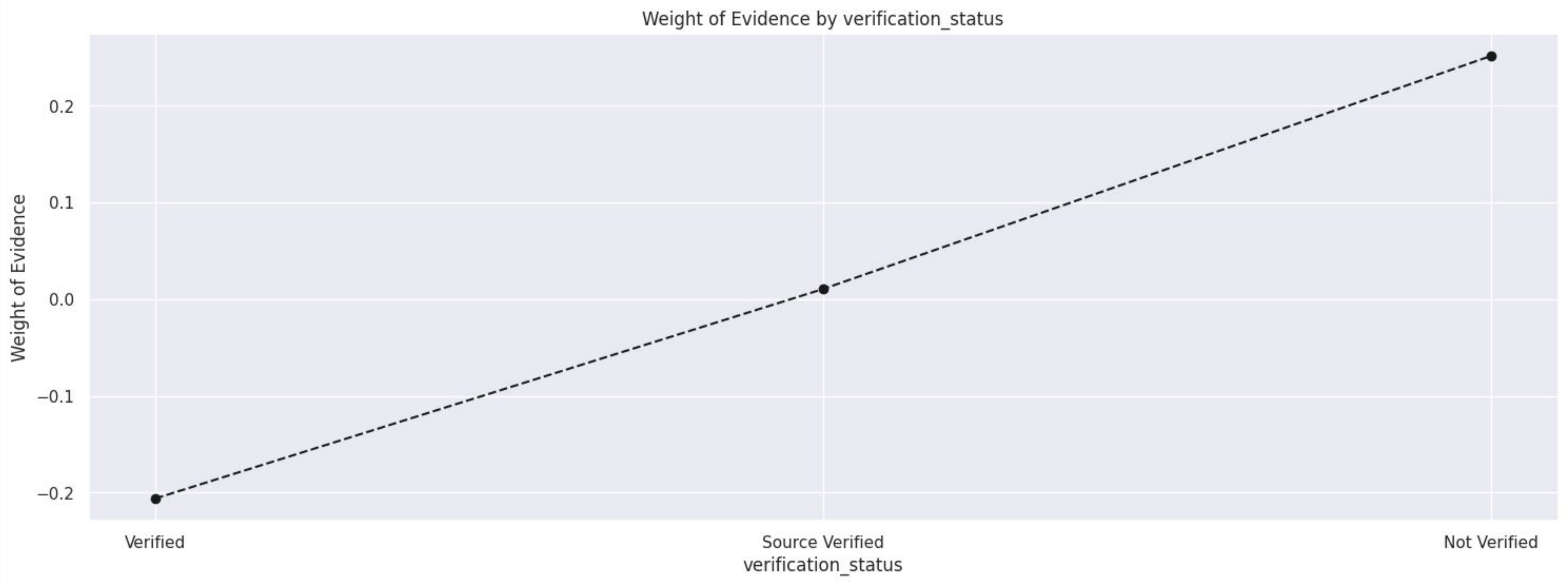
Weight of Evidence



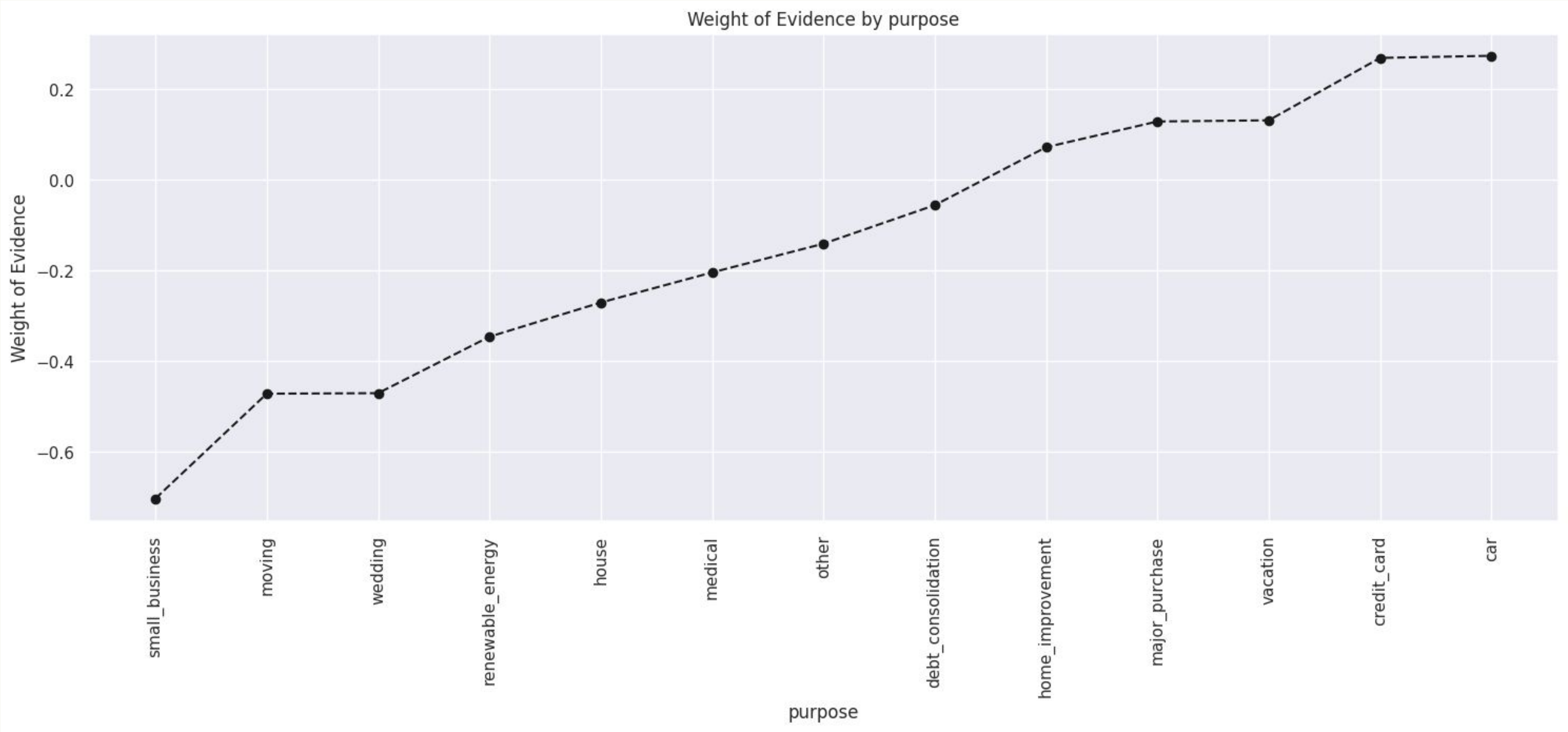
Weight of Evidence



Weight of Evidence

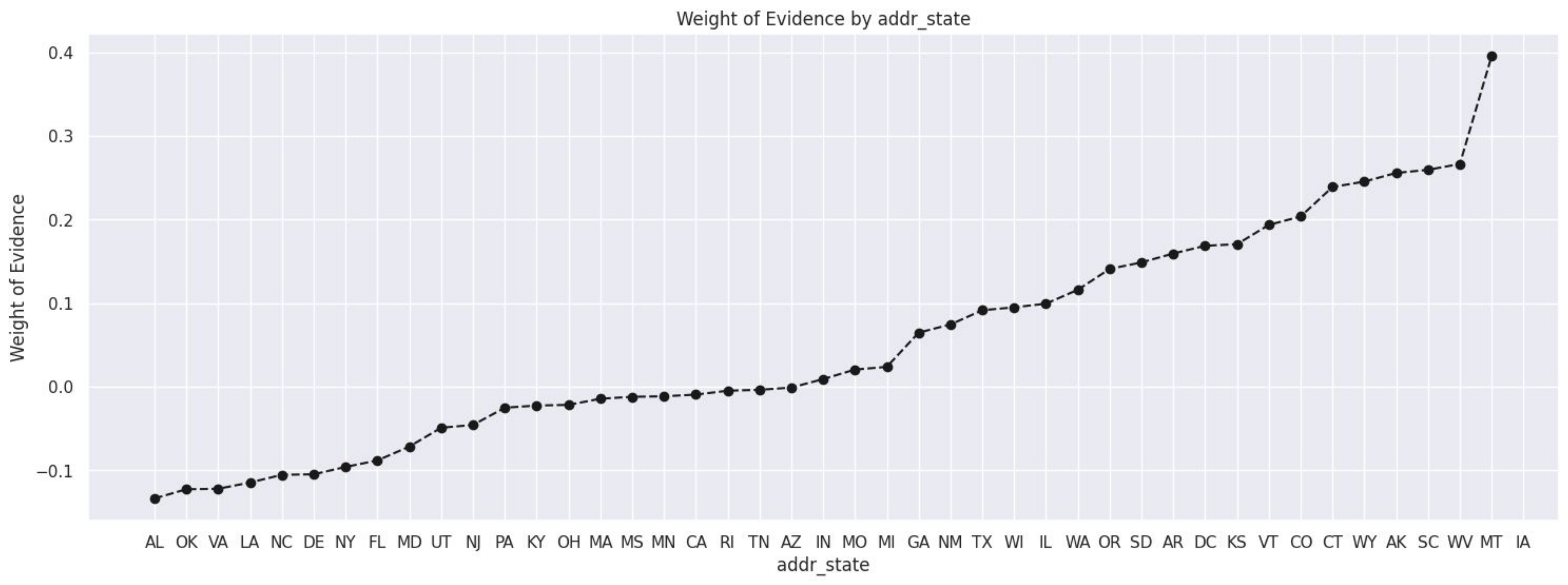


Weight of Evidence



Weight of Evidence

Weight of Evidence



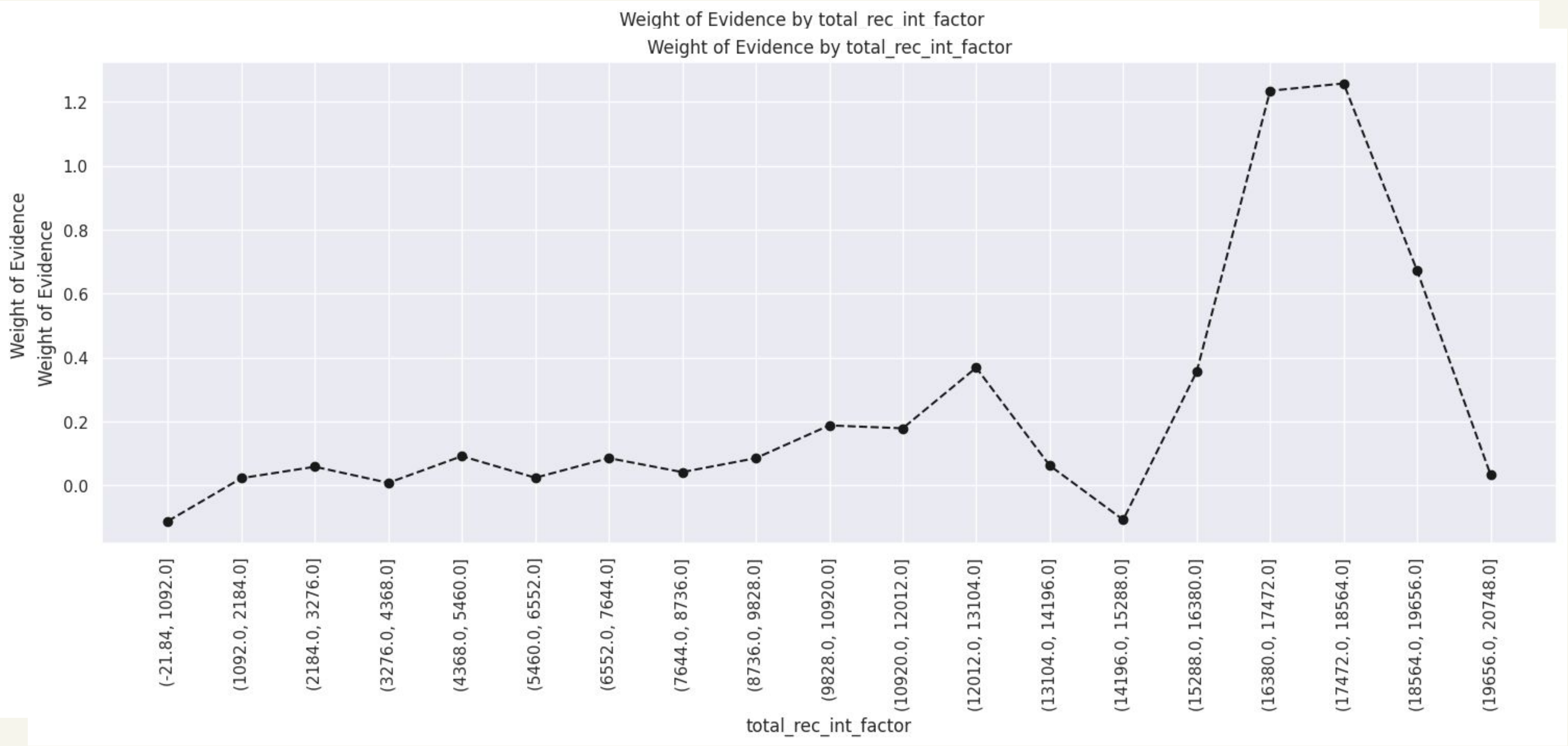
Weight of Evidence

Weight of Evidence

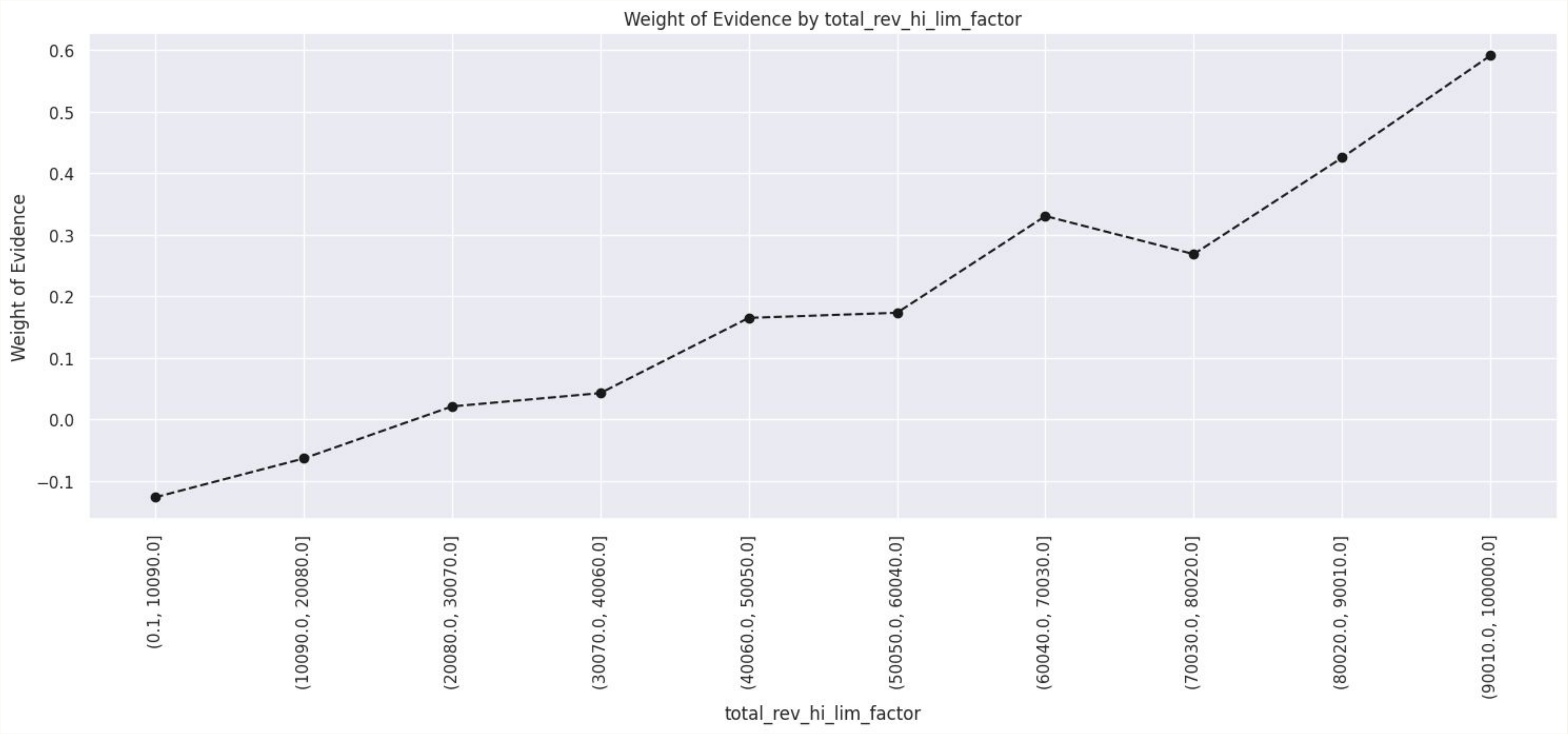
Weight of Evidence

Weight of Evidence

Weight of Evidence

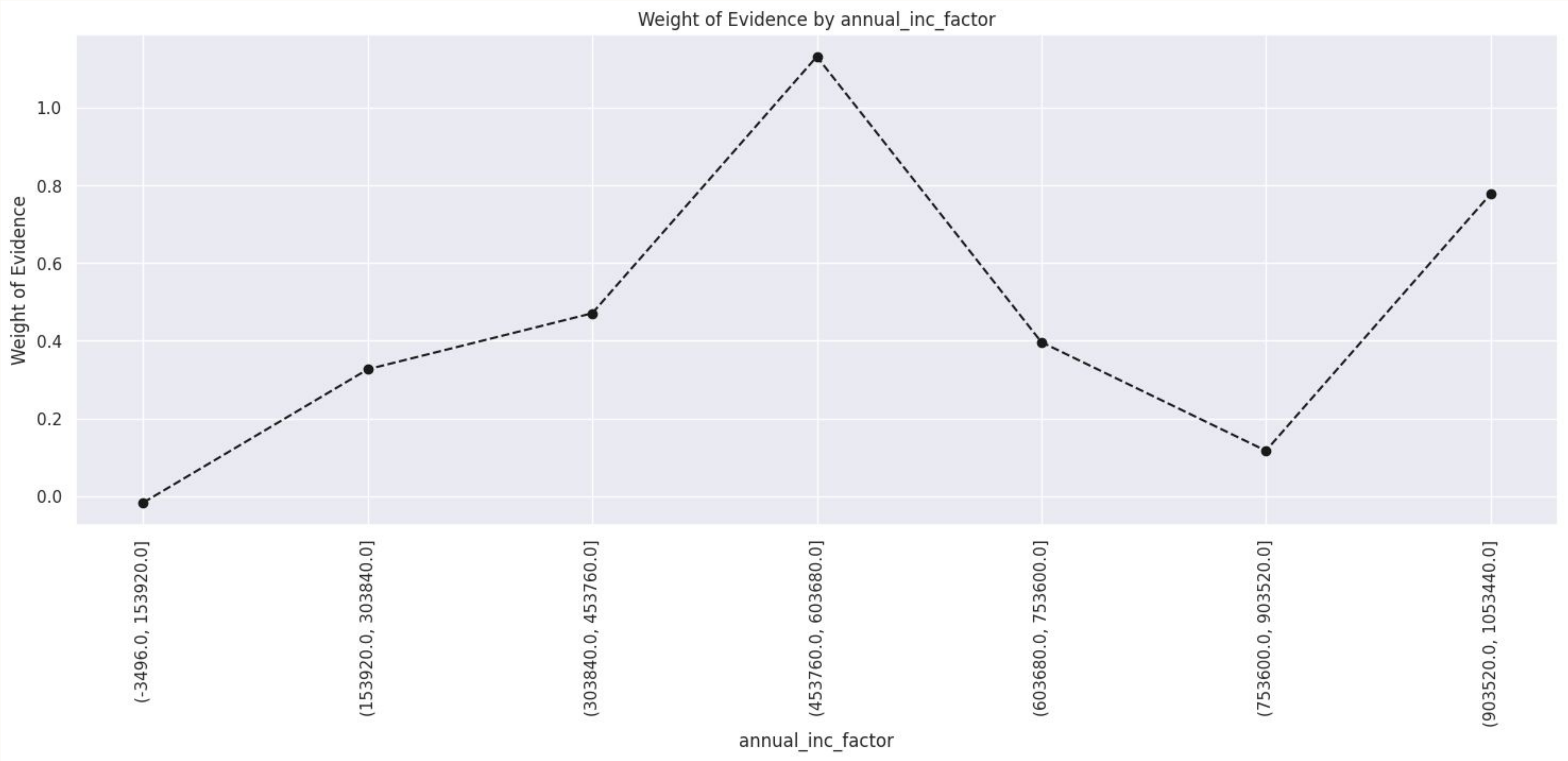


Weight of Evidence

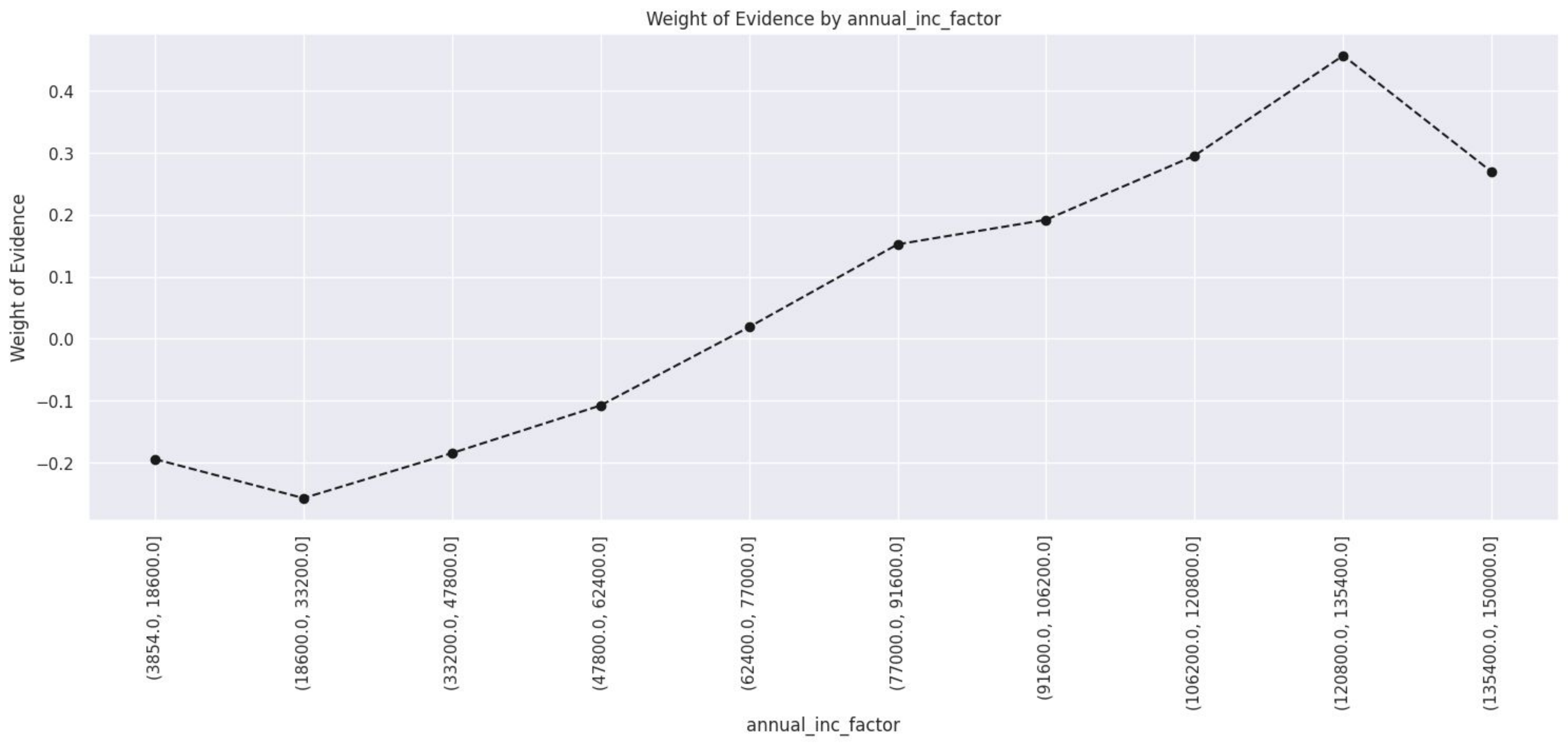


Weight of Evidence

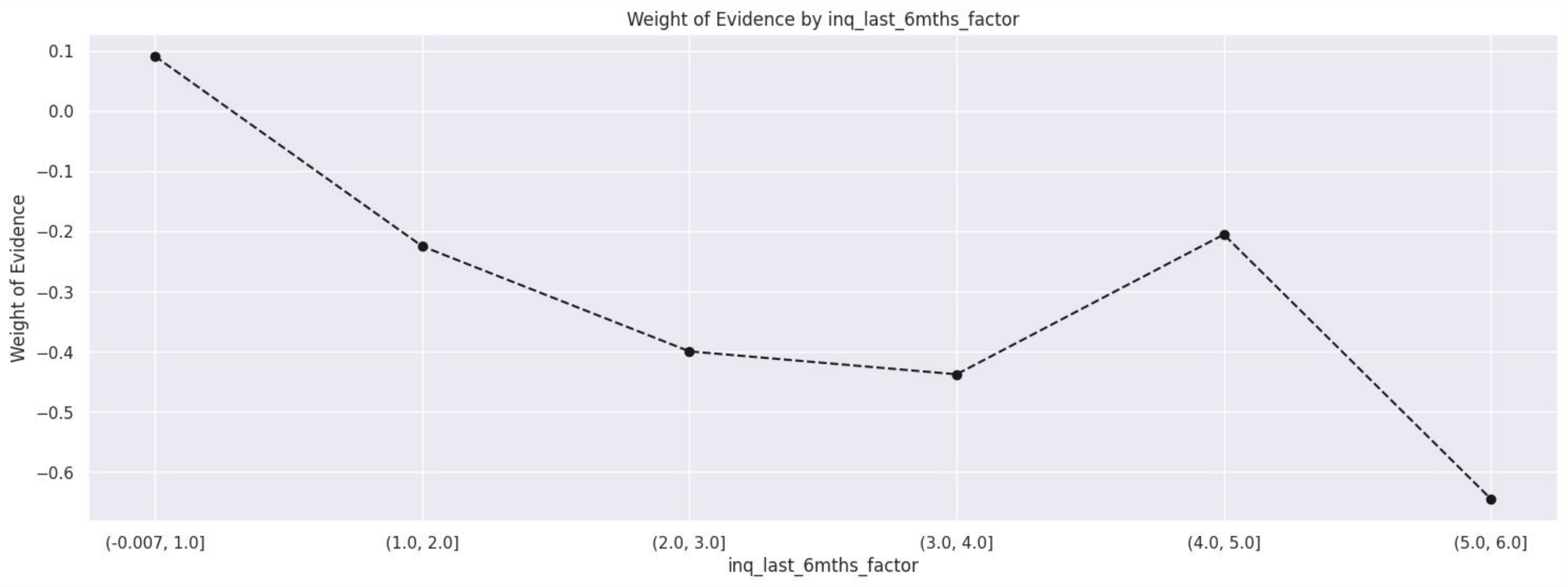
Weight of Evidence



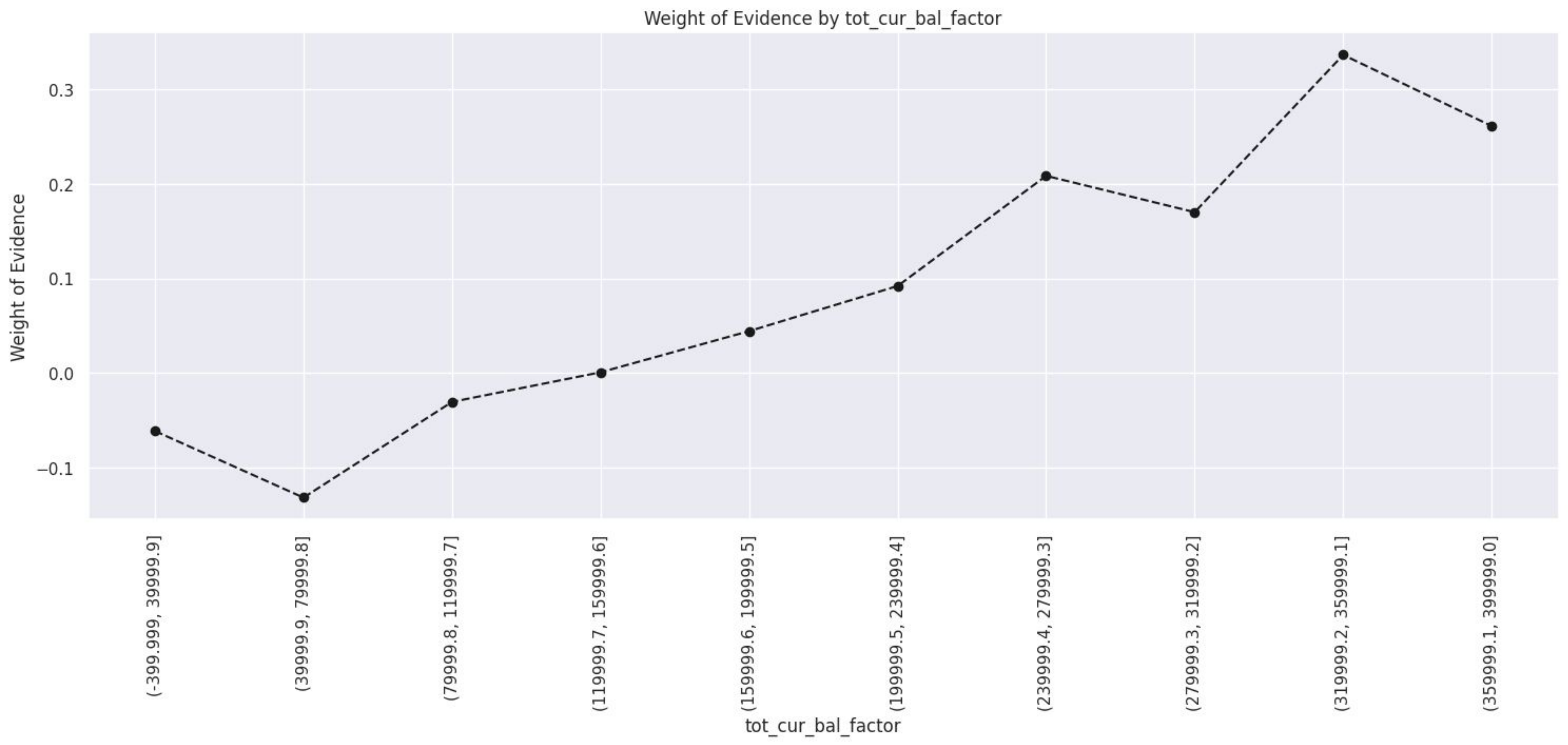
Weight of Evidence



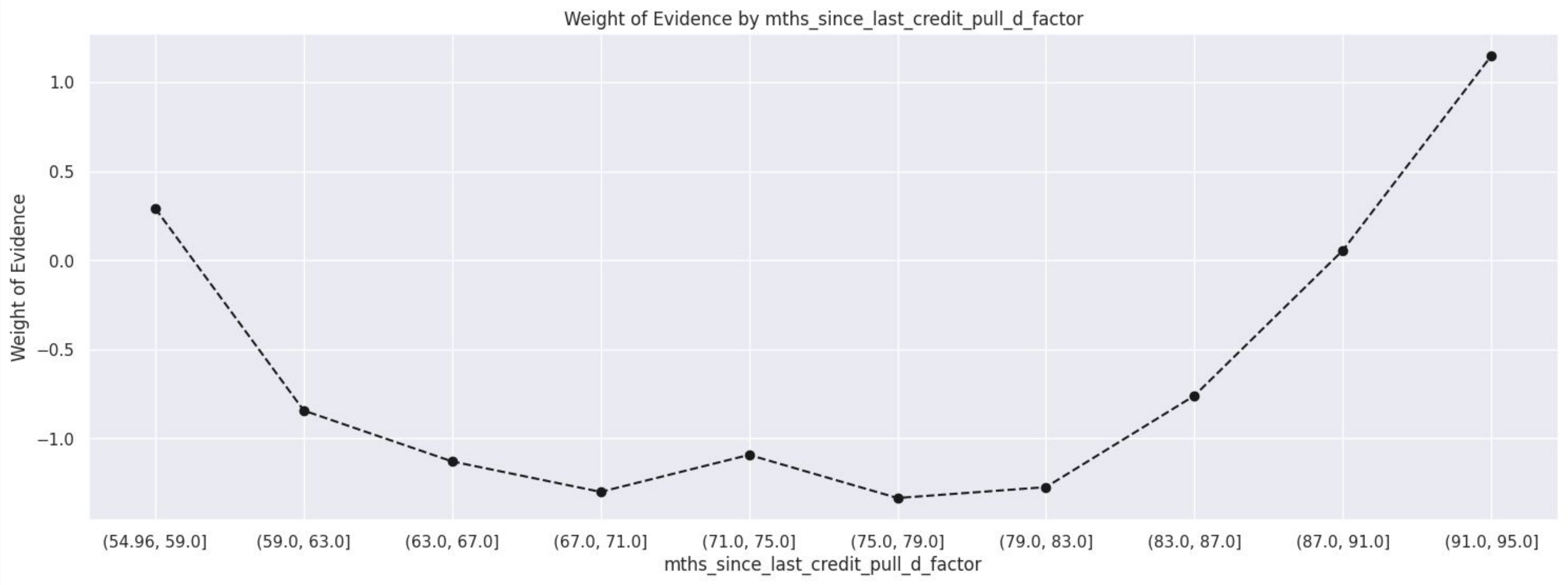
Weight of Evidence



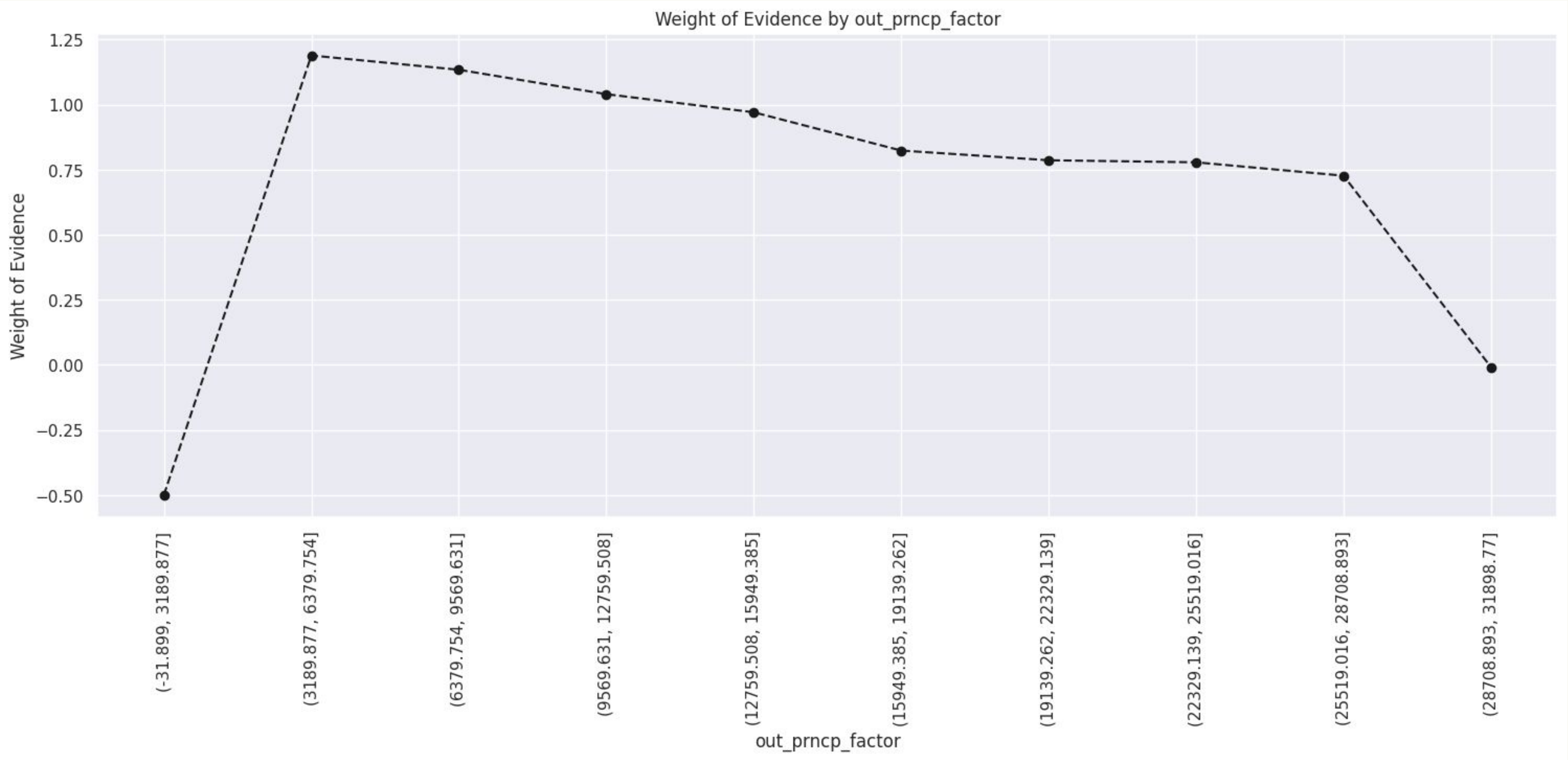
Weight of Evidence



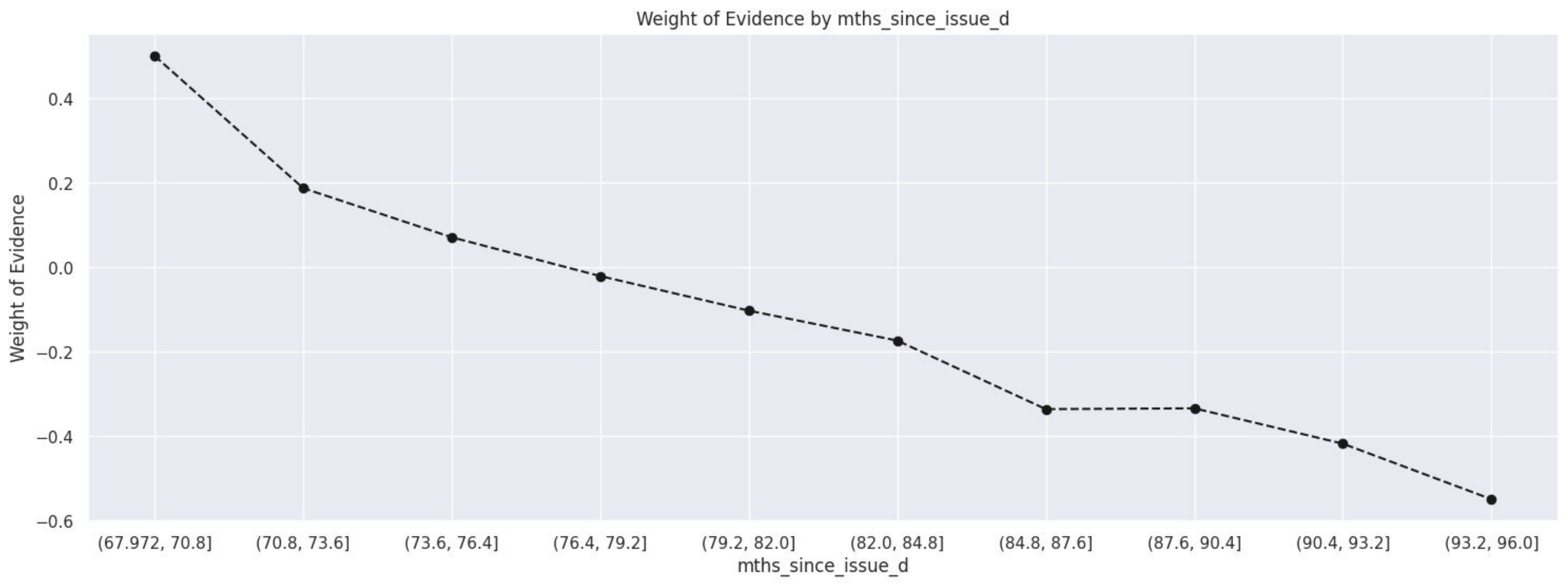
Weight of Evidence



Weight of Evidence



Weight of Evidence



Class Balance