Designing Policy Recommendations to Reduce Home Abandonment in Mexico

Klaus Ackermann
Monash University

Sue He
University of Virginia

Paul van der Boor
University of Chicago

Eduardo Blancas Reryes
Center for Data Science and Public Policy
University of Chicago

Thomas Anderson Keller
UC San Diego

Romana Khan
Northwestern University
OUTLINE

01 | Introduction
02 | Problem Formulation
03 | Data Sources
04 | Location Matching
05 | Features
06 | Modeling
07 | Results
08 | Conclusions
NO. 1

Introduction
This type of planned development sprang up on the outskirts of Mexican cities starting in the 1990s.

Millions of Mexicans got higher salaries (but still low) and government aid so that they had demand for the home.

Locals call the houses birdcages. Yet people still bought into the idea of hope, new life, and away from high criminal rate of city.
**Target**: Aimed to deal with the shoddy building, erratic urban expansion, and social justice.

**Implement**: Largely fueled these suburbs by financing mortgages with low-interest rate for Mexican workers.
Recently, there are abandoned and vandalized homes on nearly every street. People get high in these abandoned houses. They sleep in them and keep tabs on the neighbors’ comings and goings, so they can break into their homes and rob them. High abandonment => High crime rate.
NO.1 Introduction - Reason

- After building a suburb, there's often no continuing maintenance.
- Long distances to workplaces and schools.
- Lack of services, finances
- Security concern.

Residents run businesses out of their homes.
Infonavit wants to reduce home abandonment, but the current process is reactive and often too late.

- Stop making payments
- Enter a portfolio of low performing loans.
- Payment collectors repeatedly visit each home
- After extension, home marked as abandoned

12 – 18 months

- At that time, abandoned houses already motivated neighbors to leave their house behind.
- Neighborhood regeneration programs may be too late to be effective.
Reducing home abandonment in Mexico in two ways:

1. Predict the risk of home abandonment for a given individual and home information, and use that prediction to provide purchase advice to the individual.

Building a machine learning model that given an individual loan info and a home location, can predict the risk of abandonment.
2. **Provide policy recommendations** to the government using the feature of predictive model.

We can understand **maximum distance thresholds** that residents were willing to travel to services or workplaces **through data features**.
Problem Formulation

NO.2
NO.2 Problem Formulation

Binary classification problem

Problem

Outcome (a person)

Risk Prediction

Abandon

Not Abandon
This would not serve our objectives since the analysis would no longer be focused on the individual decision of a person to abandon their house. (p.06)
3.1 Loans data (Infonavit provide)

Primary data includes personal info, loan info and house characteristics for every loan granted in the last 20 years.

- **Personal info:**
  - Demographics: e.g. Age, Marital status
  - Financial information: e.g. wage, value of accumulated savings
  - Risk Index: computed by Infonavit to estimate the risk loan holder lost their job.

- **Loan info:**
  - Value of the loan
  - Interest rate
  - Value of any subsidy received

Transform the data, making it to the house level.

Exclude loans that had mismatched years or prices or not match up with other data sources.

After First Cleaning Process, 4.1 million observation at the house level.
3.1 Loans data (Infonavit provide)

Primary data includes personal info, loan info and house characteristics for every loan granted in the last 20 years.

- **House characteristics:**
  Exact location of the house is often missing. Related geo-coordinates data provided by Infonavit only available since 2008
- To deal with this constraint, we are using different approaches to match loans to colonias which is described later.
Geographic distribution of home abandonment in Mexico (p.25)
We assumed that the first interruption in the payments represented the date of home abandonment.

There were 4 types of events which were used to identify an interruption in payments:
- Date of the application for loan extension granted (59%)
- Date of the first interruption in the payment stream with continued payments after (11%)
- Date of last payment received (11%)
- Date the home is marked as abandoned but no single payment entry, date set to 2009 (5%)

For 14% of the data we could not find a pattern and excluded these observations from the study.
3.3 Housing survey data

- Infonavit also shared the results of home surveys conducted by licensed inspectors (ECUVE). The surveys measured housing quality at the municipality level.
- Factor of housing quality index:
  - Construction quality
  - Reliable water and power supply
  - Local access to schools and hospitals
  - Availability of parks and markets
Distance to employment was among the top factors that drive home abandonment. Every five years, INEGI conducts a census of businesses in Mexico, covering businesses from small grocery stores to multinational corporations. (Also includes schools and hospitals).

Information includes:
- number of employees
- owner
- address
- geolocation coordinates
3.5 Municipality data

- Number of Homicides
- Natality
- Mortality
- Natural disaster incidences
- Years of schooling (from the Population Census)
- Households statistics (from the Population Census)
- Literacy rate (from the Population Census)
- Healthcare coverage (from the Population Census)
- Number of vehicles and passenger buses
3.6 Data limitations and project scope

The loans dataset spans the last 20 years, but house coordinates were only available for loans granted after 2008. Location features were critical, so the scope of the project was limited to 2008 to 2015. A timeline of the data is presented below.
NO.4

Location Matching
Exact location of the house is often missing. Related geo-coordinates data provided by Infonavit only available since 2008.

Features like Number of schools within 5KM

After First Cleaning Process, 4.1 million observation at the house level (p.15)

Only 2 Million (49%) location info. is available

Geolocate at the colonia level by grouping all houses in the same colonia in a point.
• Match loans that had geographical coordinates with their corresponding colonia.
• Locating non-matched loans:
  Finding their closest matching-loan and assigning the same colonia, this process was limited to a distance up to 1km.
• In the end we were able to geolocate 2.4 million records (58.5% of total loans).
• P.17
After receiving and cleaning the data, we proceeded to build features.

**Location features**
The features capturing the state of the local services and infrastructure in a comparable way across Mexico.

**Personal and loan features**
Personal and loan features (both at house level) were used for characterize each colonia with colonia average, minimum and maximum.

**Merging features**
Features table in single loan-year:
- Credit-holder characteristics
- Loan characteristics
- Average personal characteristics (colonia level)
- Municipality features
- Location features
NO.6 Modeling
What is the risk of abandonment for an existing loan in the next year?

- Training a variety of machine learning models from 2008 to 2014
- Testing on loans active during 2015
- Algorithms:
  - Support Vector Machines
  - Random Forests
  - AdaBoost
  - Logistic Regression.
NO.7

RESULTS
• Best performing model: Random Forest
  - Achieving an AUC of 0.70 as in Figure

• With a 0.5 threshold, the model captured 55% of abandoned houses

• Given the inherent imbalanced nature of the data, this model produced 266,670 false positives in comparison to 4602 true abandoned houses.

• To address limited performance, we provided Infonavit with a series of data collection recommendations (see later)
Table 1: Top personal/loan predictors

<table>
<thead>
<tr>
<th>Risk Index</th>
<th>(Risk of employment loss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years since loan granted</td>
<td></td>
</tr>
<tr>
<td>Loan sales price</td>
<td></td>
</tr>
<tr>
<td>Daily wage</td>
<td></td>
</tr>
<tr>
<td>Loan account value</td>
<td></td>
</tr>
<tr>
<td>Colonia minimum sales price</td>
<td></td>
</tr>
<tr>
<td>Loan interest rate</td>
<td></td>
</tr>
<tr>
<td>Loan holder age</td>
<td></td>
</tr>
</tbody>
</table>
### NO.7 Results – Top Predictor (Government Related)

<table>
<thead>
<tr>
<th>ECUVE index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of restaurant employees within 5 km</td>
</tr>
<tr>
<td>Number of employees within 5 km</td>
</tr>
<tr>
<td>Number of businesses within 5 km</td>
</tr>
<tr>
<td>Number of hospitals within 5 km</td>
</tr>
<tr>
<td>Loan-holder age * Number of schools within 5 km</td>
</tr>
<tr>
<td>Number of churches within 5 km</td>
</tr>
<tr>
<td>Municipality Passenger buses over 1000 inhabitants</td>
</tr>
</tbody>
</table>

**Table 2: Top predictors that can be improved through policy changes**
The predictive model was deployed as a web application:
- Enter personal characteristics and a selected colonia for the home to estimate the risk of abandonment for the next year.
Conclusions
Conclusions

- Before our project, the institute had no way to estimate the risk of home abandonment at the loan level.

- Local Agency can advise Mexican workers when they are applying for a loan, so that home abandonment can be prevented.

- Improve planning for new developments.

- Prevent the spread of home abandonment in certain neighborhoods by pre-emptive action.

- Influence public policy at the federal level.
The End
THANK YOU!
Problem Formulation
NO.2 Problem Formulation

- 更换图片方法：点击图片后右键，选择“更改图片”即可。

- 标题数字等都还可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。

- 建议正文8-14号字，1.3倍字间距，标题数字等都可以通过点击和重新输入进行更改。

- 标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。

- 建议正文8-14号字，1.3倍字间距，标题数字等都可以通过点击和重新输入进行更改。
NO.3 Data Sources

A 点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可，标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。

B 点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可，标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。

C 点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可，标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。

D 点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可，标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。
点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可。标题数字等都可以通过点击和重新输入进行更改。

点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可。标题数字等都可以通过点击和重新输入进行更改。

点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可。标题数字等都可以通过点击和重新输入进行更改。

点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可。标题数字等都可以通过点击和重新输入进行更改。
NO.4 作品展示

点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可。
标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。

点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可。
标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。

点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可。
标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。
NO.4 作品展示

点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可。
标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。
NO.2 作品概述

更换图片方法：点击图片后右键，选择“更改图片”即可。标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。建议正文8-14号字，1.3倍字间距。标题数字等都可以通过点击和重新输入进行更改。标题数字等都可以通过点击和重新输入进行更改。

01 点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可。标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。

02 点击此处添加标题
更换图片方法：点击图片后右键，选择“更改图片”即可。标题数字等都可以通过点击和重新输入进行更改，顶部“开始”面板中可以对字体、字号、颜色、行距等进行修改。