# Nationwide Residential Real Estate Market Data Analysis of Negative Equity

FINAL REPORT
FHWA Task Order 693JJ319F0001
December 23, 2019
Prepared for: FHWA Office of Real Estate Services



U.S. Department of Transportation

Federal Highway Administration

# Table of Contents

Table of Contents	i
Executive Summary	1
1. Introduction	2
1.1. Project Background	2
1.2 Project Objective	3
2. Research on Potential Data Vendors	4
2.1. Data Requirements	4
2.2. Market Research on All Potential Vendors	5
2.3. Initial Evaluation of Potential Vendors	5
2.4. Further Evaluation of Vendors	6
2.5. Concluding Observations on Vendor Selection	6
2.6. Final Selection of Vendor	7
3. Data Collection, Processing, and Analysis	8
3.1. Data Procurement	8
3.2. Definition of Analysis Terms	8
3.3. Data Processing and Validation	9
3.4. Data Analysis	10
3.5. Additional QA/QC Undertaken	11
4. Results	12
4.1. Overall Results	12
4.2. State-level Dashboards	20
4.3. Comparison with Other Recent Reports and Analyses	21
5. Conclusions	23
6. Recommendations	24
6.1. Recommendations for FHWA's Waiver Program	24
6.2. Recommendations for Additional Analyses	24
Appendix: State Dashboards	A-1

# Figures

Figure 1. Dashboard of nationwide results
Figure 2. Map of U.S. States showing recovery status
Figure 3. Ranking of each State in terms of recovery percentage (current median sale prices as a percentage of pre-crash median sale prices)
Figure 4. Ranking of each State in terms percent of zip codes recovered
Figure 5. Ranking of each State in terms of pre-crash median sale prices, highlighting recovered (solid green) vs. unrecovered (hatched yellow) States
Figure 6. Explanation of the information contained in each State dashboard
Figure 7. Plot of the U.S. House Price Index, published by the St. Louis Federal Reserve Bank21
Figure 8. Plot of cumulative seasonally adjusted price change relative to the April 2007 peak for the United States, published by FHFA
Figure 9. Plot of median home sale prices since 2005, published by ATTOM Data Solutions
Figure 10. Maps of Memphis, TN, showing recovered (green) and unrecovered (purple) zip codes 25
Figure 11. 90th percentile, median, and 10th percentile sale prices for each aggregation year for the 89108 zip code in Las Vegas, NV
Tables
Table 1. Number of records received from data vendor
Table 2. Tabular results for all U.S. States, including recovery status, recovery percentage, percent of zip codes in that State that have recovered, and the pre-crash median sale price

## **Executive Summary**

This report by The Cadmus Group, Inc. is the result of a contracted study and analysis of comparative and determinate sale price data of owner-occupied single-family residential dwelling (SFRD) units to facilitate the Federal Highway Administration's (FHWA) consideration of whether there is a need to continue the Temporary Waiver of Methodology for Calculating a Replacement Housing Payment (Waiver). This study consisted of a nationwide examination of residential real estate market sale price data by gathering, documenting, compiling, and comparatively analyzing the data for each State aggregated to the zip code level. Data was gathered from 2006, prior to the historically unique 2008 real estate market crash, through the end of July 2019 (the latest reliable data available).

To initiate this analysis, market research and evaluation was undertaken to generate a list of potential real estate market data vendors that met the criteria of providing SFRD unit sale price data for the specific time period. Based on the market research and detailed evaluation of the finalist vendors, a recommendation was made for First American Data Tree (FADT) to be selected as the data vendor. It was further recommended that only one data vendor was necessary and would be required to achieve the objective of the study. Multiple purchases from different vendors would come at substantial cost, and would result in acquisition of duplicate sets of essentially the same data that would not deepen the understanding of the equity trends in the residential real estate market.

The data received from the vendor had undergone initial filtering to generate a dataset of owner-occupied residential structures that were sold as arm's length transactions under market conditions (34,513,676 transaction records). During quality control checks, a small percentage (approximately 3%) of anomalous data were excluded, representing erroneously keyed large transaction amounts (over \$40 million) and multiple parcels included in a single sale price. The data were then grouped by State, zip code, and aggregation year, and the analysis was conducted. For the zip code level analysis, summaries of zip codes with fewer than 30 transactions per year were not included; however, these transactions are included in the State and National level analyses. Additional quality assurance and quality control (QA/QC) was undertaken following the analysis.

The nationwide results indicate that the residential real estate market in the United States has returned to its pre-crash level. There are eleven states that have not yet regained 100% of their pre-crash median sale prices, with only two States that have regained less than 90% of their pre-crash median sale prices. Examined at the State and zip code level, a variety of trends are apparent in different States and in the various zip codes within each State, depending on specific conditions in each market. Several publicly available large scale or general analyses were reviewed and compared against the results of the intermediate scale analyses carried out in this project, which validated the findings.

Based on review of the sales data, a recommendation is made to allow the Waiver to expire. However, it is important to note that there may be a case by case need for the Waiver because a few markets have not yet fully recovered. A further recommendation is for additional spatial and trend analyses of the data, as such analyses would provide valuable insights to FHWA's Office of Real Estate Services that are beyond the scope of this project.

Please note that any pages that are shown partially blank are intentional.

## 1. Introduction

This section provides the background to the project as well as the overall objective of the project.

## 1.1. Project Background

Under the Uniform Relocation Assistance and Real Property Acquisition Policies Act of 1970, as amended (Uniform Act), persons displaced by a Federal financially-assisted public improvement project are entitled to specified benefits designed to ensure there are no disproportionate injuries suffered when they are required to relocate as a result of projects designed to benefit the public as a whole. One such benefit provided to homeowner-occupants is the Replacement Housing Payment (RHP) price differential, which bridges any gap between the just compensation they are constitutionally entitled to receive for the acquisition of their residential property, and the additional cost determined necessary to purchase a comparable replacement dwelling and site. In describing one of the fundamental objectives of the Uniform Act, the legislative history makes clear that a displaced homeowner should not be left worse off economically than he or she was before displacement, and should be able to relocate to a comparable dwelling that is decent, safe, and sanitary—that sentiment is codified in Title 42 of the United States Code (U.S.C.) §4621.

The historically unique global financial crisis of 2008 resulted in significant impact to the U.S. real estate market. Many State agencies are faced with the prospect of displacing homeowner-occupants who hold negative equity in what is commonly referred to as "upside-down" or "under water" mortgages, where the fair market value of the property (or just compensation) is less than the outstanding debt (the mortgage). In these circumstances, a negotiated administrative settlement to acquire the property for a federally-assisted project may eliminate the RHP otherwise available. Since the need to satisfy the mortgage balance arises as a direct result of the acquisition of the property for a project or program which is federally funded or assisted, the displaced homeowner-occupant should not be deprived of an RHP, making his or her situation worse than it was before displacement. In other words, but for the displacement caused by the transportation project, the homeowner-occupant could continue to remain in the property, and pay the periodic mortgage payments until real property values recover from the effects of the 2008 crash.

The historically unique negative equity housing situation created circumstances that justified the FHWA's 2009 Waiver. The Waiver provides a modified methodology for the calculation of an RHP under 49 CFR §24.401(b)(1).

FHWA utilized publicly available real estate market data, and general real estate market information reported by State Departments of Transportation (DOTs) and other Federal agencies to determine the need for implementation of a programmatic waiver in 2009. Since 2009, the Waiver has generally been extended in two-year increments. FHWA has continued to monitor the indicators of whether the National residential real estate market has recovered to an extent that this Waiver is no longer necessary. In publicly available data, the residential real estate market recovery rates in the various States are reported from a general aspect or from an overall rural and urban viewpoint.

Recent U.S. housing sales price data from years 2006-2017 in a CoreLogic analysis titled "U.S. Home Prices Have Returned to 2006 Peak" showed that overall National home sale prices returned to the April 2006 peak sale prices. However, the CoreLogic report also notes that while several States have exceeded peak 2006 sale prices, several others continue to have sale prices below their 2006 peak.

Additionally, Office of Real Estate Services outreach with its State and Federal partners and stakeholders for insight about local residential real estate market conditions resulted in more than 50% of the FHWA's Division offices confirming that the Waiver is still necessary in their State, and 75% of responding Federal agencies concurring the need still exists.

These factors were the catalyst for this study to be initiated, and to determine whether the need for extension of the Waiver remains, or whether it can be sunset at its current expiration date on December 31, 2019.

## 1.2 Project Objective

This project analyzed comparative and determinate data for sale prices of individual residential dwelling units, to facilitate FHWA's consideration of whether to the continue the Waiver. This nationwide examination of residential real estate markets consisted of gathering, documenting, compiling, and comparatively analyzing residential real estate sale price data in each zip code of each State. The time analyzed was from 2006 through 2019. This period covers the height of the residential real estate market in late 2006, its subsequent decline, then the historically unique 2008 crash of the nationwide real estate market, and ends with the most recent complete SFRD data available at the time the project was conducted.

<sup>&</sup>lt;sup>1</sup> U.S. Home Prices Have Returned to 2006 Peak, Home Price Index Highlights: August 2017, Molly Bosel (October 3, 2017), available at <a href="https://www.corelogic.com/blog/2017/10/homeprice-index-highlights-august-2017.aspx">https://www.corelogic.com/blog/2017/10/homeprice-index-highlights-august-2017.aspx</a>

## 2. Research on Potential Data Vendors

This section documents the research conducted to evaluate and select a vendor from which to procure the necessary real estate data to meet the project objectives. It provides information on FHWA's requirements for the data, the potential vendors and their data products/sources, and the criteria and parameters used for vendor evaluation and selection.

## 2.1. Data Requirements

At a minimum, the data needed to support the project objectives are two timeframes of parcel-level data for all 50 States, plus the District of Columbia, for sales of owner-occupied SFRD units to be purchased from private vendor(s). One dataset consists of the most recent months of available data in 2019, and the other is a pre-recession comparison dataset consisting of an identical number of months from 2006/2007, prior to the 2008 real estate market crash. The definitions of fee simple, arm's-length, and market are taken from the Dictionary of Real Estate Appraisal, 6<sup>th</sup> Edition. These definitions are as follows:

**Arm's-length transaction.** A transaction between unrelated parties who are each acting in his or

her own best interest.

**Fee simple estate.** Absolute ownership unencumbered by any other interest or estate, subject only to the limitations imposed by the governmental powers of

taxation, eminent domain, police power, and escheat.

Market.

1. A set of arrangements in which buyers and sellers are brought together through the price mechanism; the aggregate of possible

buyers and sellers and the transactions between them.

2. A gathering of people for the buying and selling of things; by extension, the people gathered for this purpose.

At a minimum, the parcel level data needed to include the following fields:

- Location Referencing System (LRS) attributes such as zip codes, county/city, geo-codable address, latitude/longitude, and/or census block information for identifying market areas.
- Sufficient data to establish (to a reasonable degree of certainty) that a market sale of each property through an arm's length transaction under market conditions has occurred. For this, deed transfer data alone is not sufficient. Sufficient data includes:
  - o A Sales Validation Questionnaire (SVQ) or Certification of Value (COV).
  - o Issuance of a mortgage by a federally regulated lending institution.
  - o Issuance of a title insurance policy by a regulated insurance institution.
  - o Registration of a new real property transfer tax.
- <u>APA's Land Based Classification Standards</u> coding was used to establish that the parcel transferred constitutes an owner-occupied SFRD unit via:
  - 1. Land area and/or existing building area so that sale price per square foot may be reliably determined to help assure that only arm's length market transactions are included in the datasets, as a sale price per square foot of \$1.00/sq. ft. for instance would identify motivations uncharacteristic to an arm's length transaction under market conditions, allowing these rows to be eliminated from the datasets.

2. Allow aggregation of data and comparison of geographical units (such as zip codes) across time.

#### 2.2. Market Research on All Potential Vendors

Market research was performed to identify potential data vendors and platforms.

## 2.3. Initial Evaluation of Potential Vendors

To achieve the project objective, the Contractor looked for a vendor with the capability of an analysis on a geographic (zip code) basis in the 50 States and the District of Columbia to include at a minimum the pre-crash (2006) time period in and the present day. Therefore, it is important to confirm that a potential vendor has nationwide data available for at least the two end points of this time period, though the entire time span is preferred. There are over 40,000 zip codes in the United States, so the resulting data files will likely run into tens of millions of records. A potential vendor must have the capability to identify and filter:

- 1. Records of owner-occupied residential real property sales only.
- 2. Records of only arm's length transactions, under market conditions.
- 3. Exclusion of transactions between immediate family, or Quit Claim deeds, such as a transfer to a trust.
- 4. A reasonably probable sale price.

#### Anticipated challenges included:

- Challenge: Non-disclosure States (such as Texas), where actual sale prices paid for real estate are
  not reported. Sources of primary data (such as face value of title insurance policies), or ways to
  directly calculate a sale price such as multiplying mortgage value and loan to value, or transfer
  tax and local (usually County) transfer tax rate are used. Use of calculated sale prices based upon
  AVM or Computer Assisted Mass Appraisal (CAMA) are avoided, as those are not considered to
  be market values.
- Challenge: Conducting an analysis over an extended period in a geographic area that almost certainly has different numbers of sales in the time period. Directly comparing 12,500 sales with a total sale price of \$125,000,000 pre-crash to 8,750 sales with a total sale price of \$112,000,000 in a given zip code is not defensible. The market would normally compare on a sale price per unit basis such as a sale price per square foot or a sale price per bedroom. However, in discussing the project requirements with the potential vendors a problem became evident. Prior to approximately 2010 it was common practice to overwrite assessor information on a given parcel each time a transaction was recorded. Thus, if a given owner-occupied SFRD sold in October 2006 and then again in December 2009, the parcel information may have changed and the historical "snapshot" of the parcel's characteristics at the time of the 2006 sale would be lost. So, it may be necessary to rely upon an average and median price per transaction within the zip code throughout the time period. The ability of the vendor to provide nationwide sales data coverage and historical data coverage becomes more important, since historical assessor data are almost certainly not available or not reliable.

• Challenge: The rigor and reliability of the data are a significant concern. The distinction between primary data and secondary data is especially important. Primary data is gathered and entered into a database by the vendor itself, as opposed to secondary data, which is collected and entered by others. If a vendor depends upon the rigor and reliability of the data in its business, errors or omissions will directly impact its bottom line, then the vendor is incentivized to ensure completeness and accuracy. If a vendor simply re-licenses and re-packages information collected by others without warranty, then it is not incentivized to the same degree. Some of this concern may be overcome if a vendor collects data from others, but then applies significant effort to validate and cross-check the data, and then own and manage the data itself.

## 2.4. Further Evaluation of Vendors

Contact with each potential vendor was made to verify the coverages, capabilities, and limitations of their data. Each vendor was vetted via a request for samples of data from Dallas County, TX. This location was chosen because Texas is a non-disclosure State and represents a very large and demographically diverse area. The data requested was intended to provide a sample of the "look and feel" of the data that the vendor would provide, and no analysis was conducted. The samples were evaluated to verify that:

- 1. The data would be furnished in a usable format. Delimited text (using either tab, comma, or pipe delimiter) is preferred because it allows direct import of sections of very large datasets into a usable, commercially available database or data warehouse platform.
- 2. The records furnished consist only of arm's length transactions of owner-occupied SFRD property sold under market conditions, and that no transactions of income producing real property, or non-arm's length transactions are included.
- 3. The records furnished do not contain duplicate records of the same transaction.
- 4. There is sufficient location information present to allow grouping of the data by geographic bounds (zip code, county, etc.).
- 5. The multiple records in the time period are laid out in a way that will enable:
  - Grouping by geography;
  - Additional computations in that group (averages, medians, etc.); and
  - Ready extraction of these summary statistics into tables for further analysis.

The Contractor conducted outreach and contact with the potential vendors to further verify data and support availability and capability, as well as refine the data requirements for this project.

## 2.5. Concluding Observations on Vendor Selection

It should be clearly understood that parcel information such as location is gained from Assessor offices. Since the objective of this project requires comparison of pre-crash data through current data in given geographic areas for all U.S. States and the District of Columbia, if data for a given zip code is not available in either the current, or the pre-crash dataset, or not at all, that zip code cannot be analyzed. For the required analysis to be conducted, vendors must provide a bulk dataset that includes relevant transaction data for the prescribed time period. Therefore, historical coverage by a vendor is limited to the number of Counties for which they have the data required and was compared as a percentage of the total number of Counties in the United States. A vendor that relies entirely upon secondary data from others is

at a disadvantage compared to a vendor that combines primary (self-generated and keyed) data with secondary data (acquired from other reliable sources).

## 2.6. Final Selection of Vendor

Based on the factors discussed above, procurement of data from FADT was recommended by the Contractor. The selection was based on the following considerations:

- 1. The First American group of companies has over 100 years of history throughout the United States in the transaction of residential real property and other related fields.
- 2. The First American group of companies include a significant number of property title and real estate mortgage offices throughout the Country, and so they represent the only one of the vendors that collect and key a significant amount of data themselves (primary data). The primary data is worthy of consideration in this analysis because FADT's reliance upon primary data is critical to its livelihood. Failure to ensure complete and accurate data will expose the vendor to risk. The Contractor has concluded this implies that FADT has a vested financial interest in assuring the accuracy and rigor of their primary data.
- 3. The Contractor concluded that FADT has direct ownership of the most nationwide data coverage and historically rigorous archives of the vendors analyzed, because of their combined use of validated primary and reliable secondary data sources.
- 4. The FADT data allows for a more straightforward filtering of records in the creation of the procured dataset. Transactions that include Notice of Default, Liquidation or Forced Sale, and other atypically motivated non-arm's length transactions can be excluded directly.
- 5. The FADT data allows for a more straightforward filtering of transactions that do not meet the definition of arm's length transactions in market conditions. Transfers of deeds to trusts or estates, and transfers between related parties can be excluded directly.
- 6. FADT offered expert support for the creation of the bulk dataset, the transfer of the bulk dataset, and in resolution of errors or other challenges following the transfer in a way that the other potential vendors do not. This allows for higher confidence in the quality of the data and subsequent analysis.

## 3. Data Collection, Processing, and Analysis

This section documents the process for procuring the data and provides information on the initial data processing and validation. It also documents the data analysis undertaken, provides definitions of important terms, and summarizes the QA/QC undertaken.

#### 3.1. Data Procurement

The data vendor made the bulk data file of approximately 461,135,487 records of real property transactions in all 50 States plus the District of Columbia. The data was comprised of owner-occupied SFRD units that sold as arm's length transactions under market conditions, and was made available to the Contractor on October 23, 2019. Not all records meet the requirements of this project. The data vendor applied filters to its total dataset that included:

- Sales recorded from January 1, 2006 through July 31, 2019.
- Sales of residential structures only (not including income producing residential structures such as apartment buildings, quad-plexes, etc.).
- Owner-occupied structures.
- NO partial interest transactions.
- NO immediate family member transactions.
- NO foreclosure or repossession transactions.
- NO transactions of less than \$1,000.

Upon application of these filters the resulting bulk dataset provided by the data vendor totaled 34,513,676 transaction records. It is important to note that some States do not require real property sale amounts to be reported, though the transaction itself is publicly recorded. Thus, in these States (Alaska, Idaho, Indiana, Louisiana, Maine, Mississippi, Missouri, Montana, New Mexico, North Dakota, Texas, Utah, and Wyoming) it is necessary to compute sale amounts based upon primary data. These sale amounts are not the result of Automated Valuation Model (AVM) or Computer Assisted Mass Appraisal (CAMA) estimates, though they are calculated (transfer tax / local tax rate; or, mortgage amount x loan-to-value). In non-reporting States, these calculation methods represent the best indication of sale price available, and were included in this analysis.

## 3.2. Definition of Analysis Terms

There were several terms, previously identified and defined in the market research phase, whose meanings are carried through the remainder of this report (such as arm's length transaction and market). During data processing, a variety of new terms were defined, as follows:

10th Percentile Sale Price. This is the amount (in dollars) which 10% of the sale prices are at or

below (least expensive). This is analyzed at the zip code, State, and

National levels.

90th Percentile Sale Price. This is the amount (in dollars) which 90% of the sale prices are at or

below (more expensive). This is analyzed at the zip code, State, and

National levels.

Median Sale Price. This is the amount (in dollars) that represents the 50<sup>th</sup> percentile, or the

amount at which 50% of the sale prices are at or below. This is

analyzed at the zip code, State, and National levels.

**Aggregation Year.** Each aggregation year is comprised of sales from August 1 through

July 31 of the following year. This period was chosen to allow inclusion of the most recent available data (July 2019), while also ensuring comparison of consistent months between the years. Data was analyzed for thirteen (13) distinct aggregation years, covering the entire

period from August 1, 2006 through July 31, 2019.

**Pre-crash.** The period defined as pre-crash for the purposes of this report is

comprised of sales from August 1, 2006 through July 31, 2007.

**Current.** The period defined as current for the purposes of this report is

comprised of sales from August 1, 2018 through July 31, 2019.

**Recovery Percentage.** This term is the median sale price in the current period as a percentage

of the median sale price in the pre-crash period. This is analyzed at the

zip code, State, and National levels.

**Recovered.** For purposes of this report, a market is "recovered" if it has regained at

least 100% of its pre-crash median sale price.

**Percent of Zip Codes** 

Recovered.

The percentage of zip codes in the geographic region that have recovered at least 100% of their pre-crash median sale prices. This is

analyzed at the State and National levels.

## 3.3. Data Processing and Validation

Once the Contractor imported the bulk dataset and defined key terms, sales were grouped by State, zip code, and aggregation year, and quality control checks were conducted to identify anomalies. The most significant anomalies are grouped into two categories:

- Sale prices of greater than \$40 million.
- Multiple parcels included in sale prices.

In addition to these, the Contractor identified an additional pattern in the data that needed to be addressed:

• Zip codes with few (1 to 30) transactions in an aggregation year.

After consulting with the data vendor, the vendor did additional analysis of records with sales over \$40 million and a single parcel in the sale price. Based on manual and automated verification steps, the vendor verified its best assessment that sale prices for those records were likely erroneously keyed and represent a small number of outliers in the data. The data vendor recommended that these be eliminated from consideration, and they were removed from the aggregated analysis dataset.

Data that indicated multiple parcels were included in the sale prices did not impact a large number of records. Given that these data represent a small number of transactions and that there was no reliable way

to address these sales (beyond, for example, computing an average price per property), they were eliminated from the aggregated analysis dataset.

The anomalous data that was excluded (erroneously keyed large transaction amounts and multiple parcels included in sale prices) represented approximately 3% of the dataset. Once removed, 33,318,592 filtered records remained.

The minimum number of transactions in a zip code is subtler and is a function of best practice in statistical analysis. One of the most frequent problems in statistical analysis is the determination of the appropriate sample size, which is required for validity. If the sample size it too small, it will not yield valid results. In a complex market analysis such as the real estate market, the results from the small sample size would be questionable if the presence or absence of a single record measurably effects the results. A sample size of fewer than 30 raises this concern. According to Statistics Solutions, an online forum that supports dissertation and advanced research, there is no certain rule to determine the sample size. Statistics Solutions do report, however, that some researchers support a general rule when using the sample size. For example, in regression analysis, many researchers say that there should be at least 10 observations per variable. However, in a complex analysis like that of the real estate market where there are multiple independent variables, then a more defensible rule would be to have a minimum sample size of 30. There are some advanced statistical analyses of product markets or medical research for which researchers will apply a separate statistical test to determine sample size, but these are very specific to the type of analysis being done, and documentation of a specific sample size computation for residential real estate markets was not evident in the literature. Thus, in the absence of a supportable method to quantify a specific sample size, the largest qualitative sample size found in the literature (30) was selected. In the case of the residential real property market there are many zip codes that are almost entirely agricultural, or public lands, or commercial/industrial, and so the number of SFRD arm's length transactions under market conditions is very small. In such areas, the presence or absence of even one transaction could measurably alter the analysis, and result in unreliable or unsupportable conclusions. For these reasons, summary statistics for zip codes containing fewer than 30 records were not included in the aggregated analysis dataset (although those were included in summary statistics at the State and National levels).

## 3.4. Data Analysis

After addressing the data anomalies noted above, the data were grouped by State, zip code, and aggregation year, and the following primary analyses were conducted:

- 1. Count of sales for each aggregation year from pre-crash to current.
- 2. Computation of 10<sup>th</sup> percentile sale price for each aggregation year from pre-crash to current.
- 3. Computation of median sale price for each aggregation year from pre-crash to current.
- 4. Computation of 90<sup>th</sup> percentile sale price for each aggregation year from pre-crash to current.
- 5. Comparison of pre-crash and current median sale prices to determine the current sale price as a percentage of pre-crash price.
- 6. Determination of whether the current price has recovered to 100% or more of its pre-crash price.

## 3.5. Additional QA/QC Undertaken

As discussed in Section 3.1, the QA/QC effort for this project began with the data vendor applying filters to more than 460 million records in its possession to return a dataset that is comprised exclusively of arm's length transactions under market conditions of owner-occupied SFRD properties. This resulted in a raw dataset as shown in 1, below.

Table 1. Number of records received from data vendor

<b>Total Records in Dataset</b>	<b>Total Records Received</b>	Percent Records Sent
461,135,487	34,513,676	7.48%

Once the raw dataset was received, the Contractor undertook quality control checks to identify anomalies and confirm the vendor's filters. Checks for completeness included assessing coverage of States, years, and day of the year. The QA also included a check on records of high and low value sale prices and transaction type. As expected, the distribution of records per year and records per State mirrored known market trends.

At the conclusion of the analysis to group and aggregate the data, the Contractor performed a series of QA/QC checks on the calculations. This was conducted by calculating summary statistics using two different platforms and programming languages. Both sets of calculations were conducted on an Apache Spark Databricks platform, a distributed computation framework appropriate for analysis of the approximately 34 million rows of raw data. Data were prepared using the Python programming language. Two sets of summary statistics were calculated separately using the R programming language and the Python programming language and the results were compared to confirm accuracy. These statistics included counts, median, percentiles, mean, extremes, grouped by zip code, State, and Nationally. Finally, once the charts and data visualizations were created, QA checks were performed to ensure the visuals reflected the calculated values.

## 4. Results

This section presents the results of the data analysis, including nationwide, State-level, and zip code level perspectives. It also provides a comparison against other recent reports and analyses.

#### 4.1. Overall Results

The results of the analysis performed using the data received from FADT are presented in the figures and tables that follow, along with accompanying discussion. Additionally, the Contractor reviewed and compared against other analyses in publicly available sources such as the Federal Housing Finance Agency (FHFA) as well as published reports by market participants such as CoreLogic and ATTOM Data. These sources are discussed in more detail in Section 4.3, later in this report. Based on the analysis conducted for this project, and the comparison against the other sources identified above, the United States real estate market has recovered from the 2008 crash from the perspective of median sale price of owner occupied single family residential real property.

As shown in Figure 1, the current nationwide median sale price is 116% of the pre-crash median, meaning that the Nation as a whole has recovered from the 2008 crash. It is noted that volatility in sale price is most evident in the 90<sup>th</sup> percentile sale price (the more expensive homes), and that these have recovered 103% of their pre-crash number as compared to 116% for median and 145% for the 10<sup>th</sup> percentile sale price. Both median and 10<sup>th</sup> percentile sale prices were less affected by the crash, and they have regained more of their pre-crash prices. On a State-by-State basis, the analysis indicates that forty States (including the District of Columbia) have recovered to at least their median pre-crash residential sale prices, and eleven States have not. At the zip code level (for the zip codes that were analyzed), 74% of zip codes across the United States have regained their pre-crash median sale prices, while 26% have not. A further examination of these trends is presented in more detail below. Finally, it is interesting to note that the number of arm's length transactions under market conditions has increased from 2,737,139 transactions in the pre-crash period to 2,815,444 transactions in the current period, an increase of 2.86%.

Table 2 presents the recovery status of each State, based on the percentage of the median sale prices in the current period as compared to the median sale prices in the pre-crash period (greater than or equal to 100% is recovered). The recovery percentage is shown in the third column. The fourth column presents the percent of the zip codes in the State that have recovered, and the fifth and final column presents the pre-crash median sale prices, for reference. These data are displayed in map and graphical formats in the figures that follow.

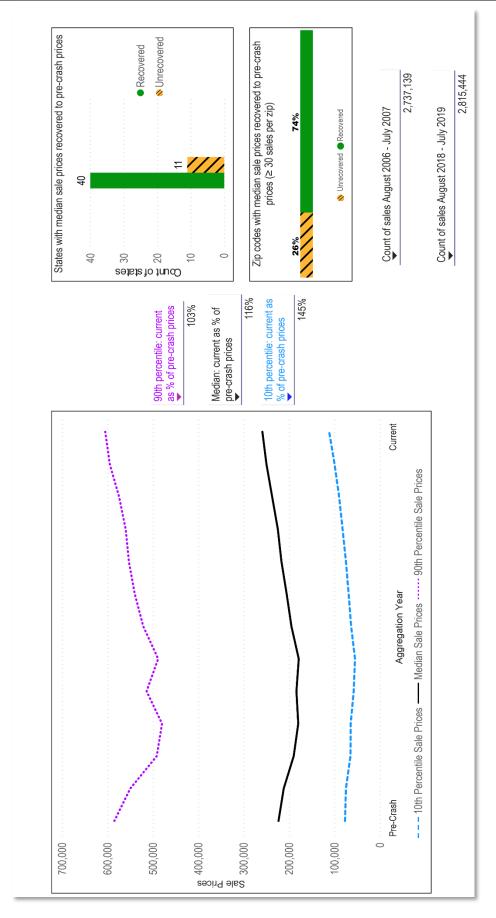


Figure 1. Dashboard of nationwide results

Table 2. Tabular results for all U.S. States, including recovery status, recovery percentage, percent of zip codes in that State that have recovered, and the pre-crash median sale price

State	Recovered?	Recovery Percentage	Percent of Zip Codes Recovered	Pre-crash Median Sale Price
Alabama	Yes	122.9%	82.1%	\$140,304
Alaska	Yes	123.1%	96.3%	\$253,950
Arizona	No	98.8%	43.1%	\$266,000
Arkansas	Yes	132.2%	89.4%	\$121,000
California	No	98.9%	65.2%	\$500,500
Colorado	Yes	152.7%	98.6%	\$238,000
Connecticut	No	87.8%	13.0%	\$279,000
Delaware	Yes	101.7%	56.7%	\$234,900
District of Columbia	Yes	126.6%	100.0%	\$409,900
Florida	No	98.0%	45.2%	\$245,000
Georgia	Yes	115.7%	85.3%	\$185,800
Hawaii	Yes	136.6%	86.5%	\$410,000
Idaho	Yes	146.0%	100.0%	\$182,800
Illinois	No	90.9%	39.2%	\$242,000
Indiana	Yes	124.3%	93.6%	\$144,000
Iowa	Yes	136.5%	97.0%	\$126,000
Kansas	Yes	146.0%	96.7%	\$147,300
Kentucky	Yes	130.2%	96.3%	\$133,285
Louisiana	Yes	130.9%	97.5%	\$149,000
Maine	Yes	112.3%	89.7%	\$217,200
Maryland	No	94.5%	34.3%	\$359,900
Massachusetts	Yes	126.7%	94.3%	\$318,000
Michigan	Yes	123.7%	70.6%	\$142,290
Minnesota	Yes	120.3%	93.5%	\$207,900
Mississippi	Yes	111.0%	85.7%	\$172,500
Missouri	Yes	118.0%	85.3%	\$158,900
Montana	Yes	137.6%	100.0%	\$189,000
Nebraska	Yes	145.6%	98.1%	\$125,000
Nevada	No	94.1%	38.7%	\$329,332
New Hampshire	Yes	104.1%	73.3%	\$255,000
New Jersey	No	90.6%	24.2%	\$358,777
New Mexico	Yes	106.0%	72.3%	\$198,100
New York	No	88.3%	60.5%	\$300,000
North Carolina	Yes	122.0%	92.5%	\$184,500
North Dakota	Yes	164.2%	100.0%	\$140,024
Ohio	Yes	106.2%	80.4%	\$141,200
Oklahoma	Yes	144.7%	97.8%	\$114,000
Oregon	Yes	132.7%	96.5%	\$259,900
Pennsylvania	Yes	108.8%	76.5%	\$188,500
Rhode Island	No	99.8%	49.1%	\$265,000
South Carolina	Yes	122.9%	87.9%	\$168,000
South Dakota	Yes	178.3%	100.0%	\$115,000
Tennessee	Yes	149.6%	97.3%	\$142,006
Texas	Yes	153.8%	98.2%	\$163,700
Utah	Yes	140.5%	98.0%	\$231,300
Vermont	Yes	121.3%	71.4%	\$200,000
Virginia	No	96.7%	64.1%	\$300,000
Washington	Yes	122.0%	99.3%	\$309,950

State	Recovered?	Recovery Percentage	Percent of Zip Codes Recovered	Pre-crash Median Sale Price
West Virginia	Yes	165.4%	81.4%	\$101,000
Wisconsin	Yes	115.9%	90.5%	\$164,000
Wyoming	Yes	126.9%	94.1%	\$197,050

When displayed geographically in Figure 2, it becomes apparent that the eleven unrecovered States are predominately coastal. Seven of the eleven unrecovered States are located on the eastern seaboard. Arizona, Illinois, and Nevada are the three unrecovered States that are not coastal. The greatest concentrations of unrecovered States are found either in the southwest or northeast parts of the Country.

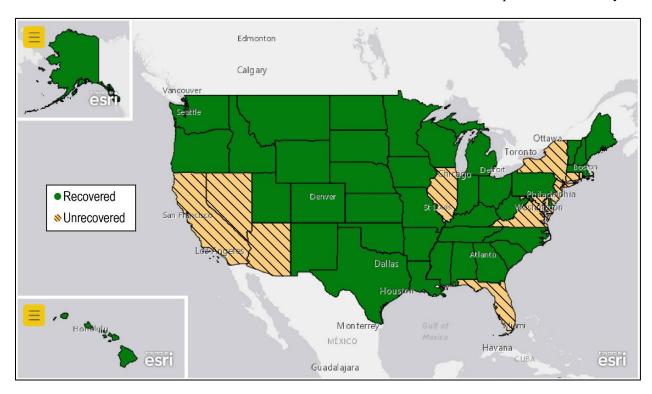


Figure 2. Map of U.S. States showing recovery status

Figure 3 illustrates the ranking of each State, in terms of recovery percentage. As previously mentioned, eleven States have not recovered to at least 100% of their pre-crash median sale prices. Of those, only two States (New York at 88.3% and Connecticut at 87.8%) have not recovered at least 90% of their pre-crash prices. Four States (Maryland, Nevada, Illinois, and New Jersey) have recovery percentages between 90-95%. Six States (Rhode Island, California, Arizona, Florida, and Virginia) have recovery percentages above 95% of their pre-crash price, but are not yet above 100%. At the other end of the ranking, five States (South Dakota, West Virginia, North Dakota, Texas, and Colorado) have recovery percentages at or above 150%.

When viewed by zip code (Figure 4), it is apparent that of the States that are not recovered from the statewide perspective, most also show a majority of their zip codes as unrecovered (eight of the eleven unrecovered States have fewer than half of their zip codes as recovered, with the remaining three unrecovered States having greater than half recovered (California, Virginia, and New York). New York is

of interest because its statewide recovery percentage is quite low (88.3%), but it's percent of zip codes recovered is quite high compared to other unrecovered States (60.5%). The driver for New York State not yet recovering may be the large number of unrecovered zips codes representing relatively high-density population areas in the suburbs and commuter areas of New York City (as visible in the zip code maps). These zip codes may have experienced more pre-crash market appreciation than other zip codes, and are significant enough to affect the entire State.

Figure 5 clearly shows that the unrecovered States have some of the highest pre-crash median sale prices. Of the eighteen States with the highest pre-crash median sale prices, eleven of them are unrecovered. This may be because the unrecovered States experienced more pre-crash market appreciation overall than other States.

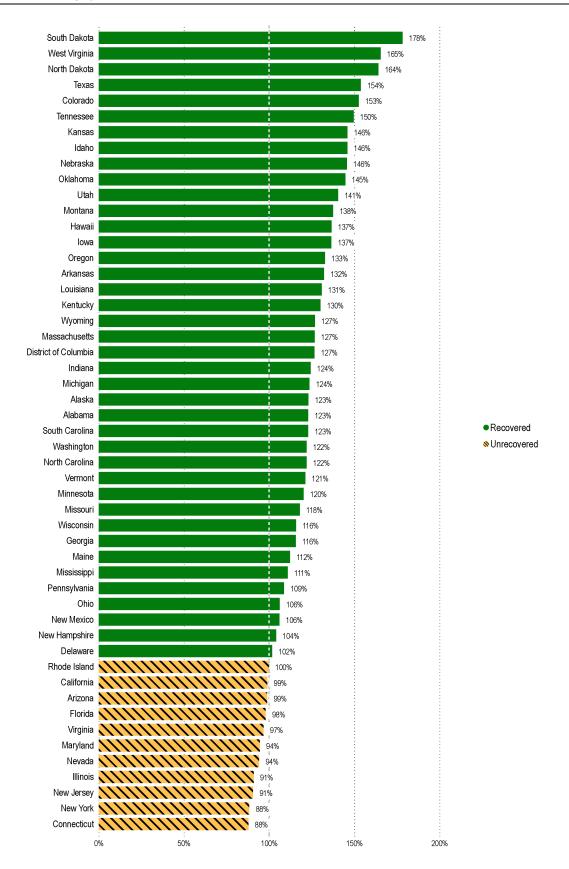


Figure 3. Ranking of each State in terms of recovery percentage (current median sale prices as a percentage of pre-crash median sale prices)

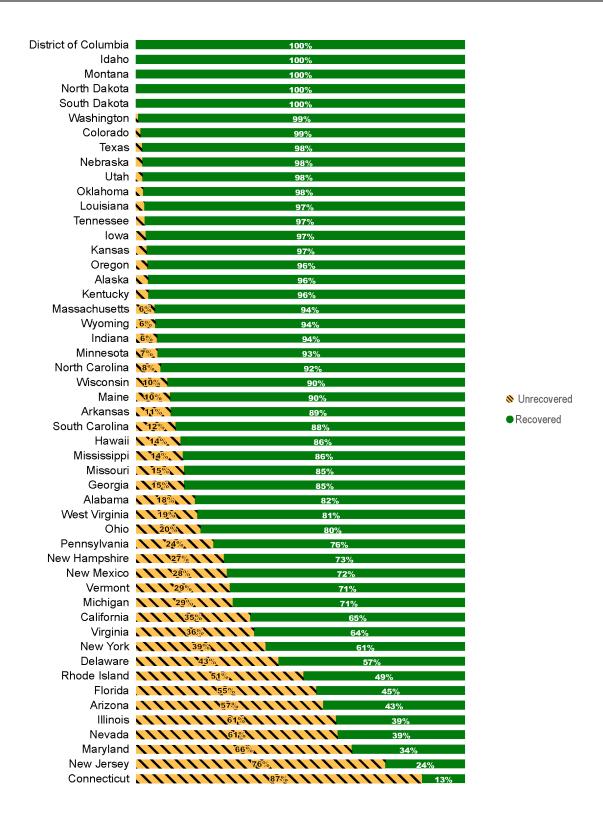


Figure 4. Ranking of each State in terms percent of zip codes recovered

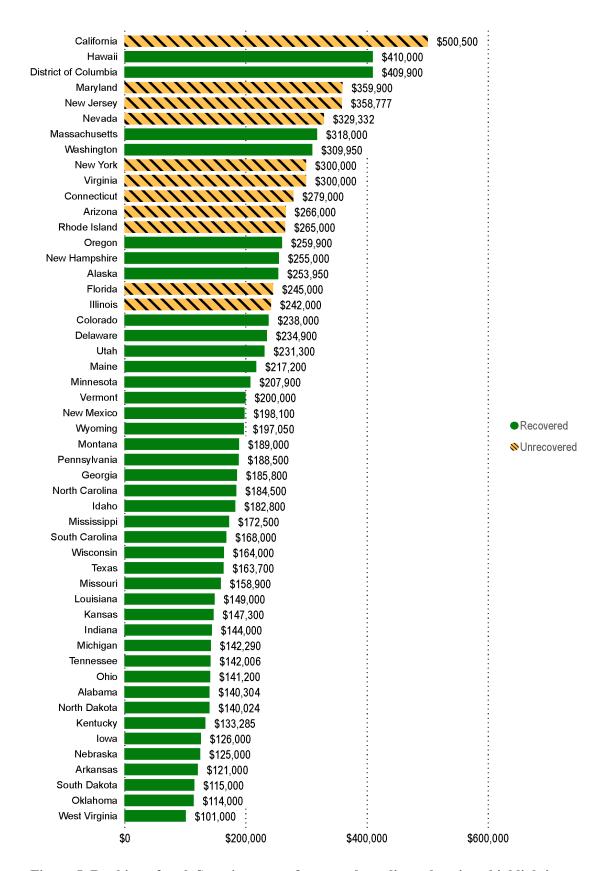


Figure 5. Ranking of each State in terms of pre-crash median sale prices, highlighting recovered (solid green) vs. unrecovered (hatched yellow) States

#### 4.2. State-level Dashboards

State-level dashboards, including recovery percentage trendlines, overall State recovery status, percent of zip codes recovered, count of sales, and maps showing the recovery status of each zip code in the State, are presented alphabetically in the appendix. The information presented in the dashboard views are explained in Figure 6.

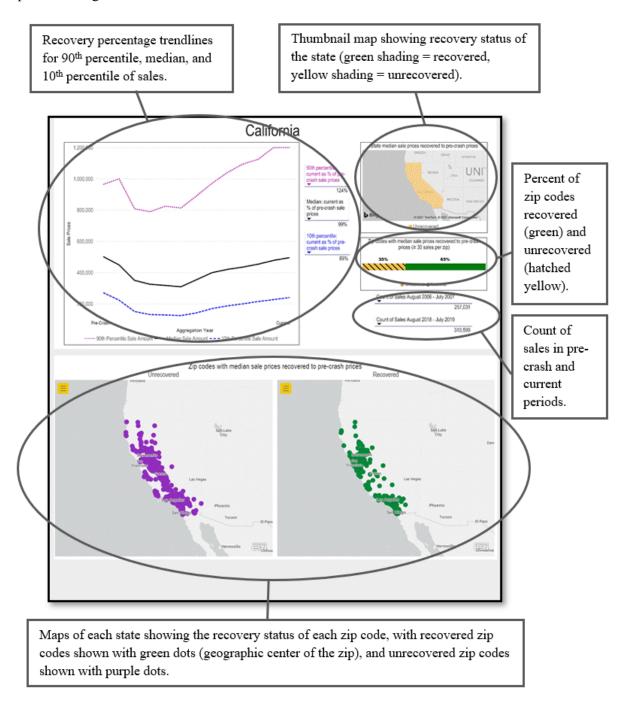


Figure 6. Explanation of the information contained in each State dashboard

## 4.3. Comparison with Other Recent Reports and Analyses

In order to compare results to other sources, publicly available market reports that have examined this issue at a larger scale and presented comparable data were reviewed. These other data, such as a plot of the U.S. House Price Index published by the St. Louis Federal Reserve Bank (Figure 7), found, like this project, that house prices "bottomed out" in 2010 to 2012 following the recession, and have since been recovering, exceeding their pre-crash levels.



Figure 7. Plot of the U.S. House Price Index, published by the St. Louis Federal Reserve Bank

The Federal Housing Finance Agency (FHFA) has published similar data in the form of seasonally adjusted House Price Index (HPI), a dataset that is limited to purchases only (Figure 8).

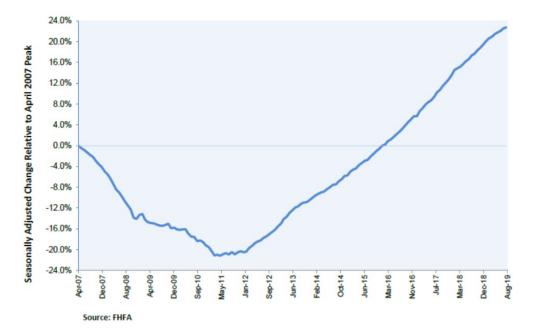


Figure 8. Plot of cumulative seasonally adjusted price change relative to the April 2007 peak for the United States, published by FHFA

Finally, a report on historical median home sale prices published by ATTOM Data Solutions in October 2019 was reviewed (see plot in Figure 9). As the other publicly available sources of information have concluded, the ATTOM report also concludes that median home sale prices have returned to or exceeded their pre-crash levels.



Figure 9. Plot of median home sale prices since 2005, published by ATTOM Data Solutions

## 5. Conclusions

This section provides the Contractor's overall analysis conclusions about trends in the National residential real estate market since 2006, and the progression of decline and recovery in each market, in order to inform the continuing need for FHWA's Waiver.

The results of the data analysis indicate that the effects of the historically unique 2008 real estate market crash on the residential real estate market varied somewhat from State to State, but generally bottomed out from 2010 through 2012, at which time most markets began to recover. There are only two States (Connecticut and New York) that have not recovered at least 90% of their pre-crash median owner occupied SFRD unit price. With 100% recovery as the benchmark, the list grows to eleven States, which are for the most part, coastal States. The 90<sup>th</sup> percentile sale prices have shown more volatility than the median or 10<sup>th</sup> percentile sale prices. The primary conclusion from this study and analysis is that generally the nationwide residential real estate market has recovered from the 2008 nationwide real estate market crash in terms of median sale prices of owner-occupied SFRD units. This conclusion also allows that there may be a need for the Waiver to remain in certain States overall, and in certain zip codes.

## 6. Recommendations

This section provides the Contractor's recommendations for implementation of the findings, in terms of FHWA's consideration of the continuing need for the Waiver and additional analyses.

## 6.1. Recommendations for FHWA's Waiver Program

Based upon the analyses contained in this report, and upon the other large scale analyses identified, the Contractor recommends the Waiver be allowed to expire. If it is to continue, it could be limited to only the States of Connecticut and New York, as these States still show recovery of less than 90% of their precrash median sale prices. FHWA may instead wish to consider continuing the Waiver on a more localized basis (such as for certain specific metropolitan areas), based on the zip code level results of this analysis.

## 6.2. Recommendations for Additional Analyses

The Contractor recommends that FHWA consider additional intermediate scale analyses on the aggregated data provided in this project that are beyond the scope of this project. There are many related analyses that could be conducted using the data acquired for this project. Examples of possible additional analyses are presented below.

Analyze statewide sale price trends over time: The eleven States that have not recovered 100% of their pre-crash median sale prices could be analyzed in more detail to determine whether they are trending up to pre-crash levels, or have plateaued at their current levels. For example, Connecticut does not appear to have trended up to any significant degree post-crash, and Illinois appears to have plateaued in recent years. Both of these States may have found a new market equilibrium point.

Examine the local market recovery status in States that have not recovered: For example, in the cases of Illinois and Nevada, the greatest number of zip codes failing to recover are found within single metropolitan areas (Chicago and Las Vegas, respectively).

Examine the local market recovery status in States that have recovered: One avenue here would be to explore what trends are preventing further SFRD unit recovery in States that have recovered. For example, recovered States such as Georgia, Michigan, Ohio, and Wisconsin have clusters of zip codes (mostly in the urbanized areas) that have not recovered. Areas such as Atlanta, GA; Detroit, MI; Cleveland, OH; and Milwaukee, WI, were hit hard by the crash and in its aftermath, and there are unmistakable clusters of zip codes in these areas that have not recovered their pre-crash SFRD unit median prices, even though there are also clusters of zip codes within those same metropolitan areas that have recovered. Another avenue of example geographic analysis is the metropolitan area of Memphis, TN, which has the largest concentration of unrecovered zip codes in the State of Tennessee, but these unrecovered zip codes are interspersed with zip codes that have recovered (Figure 10), and the State of Tennessee as a whole has experienced robust recovery.

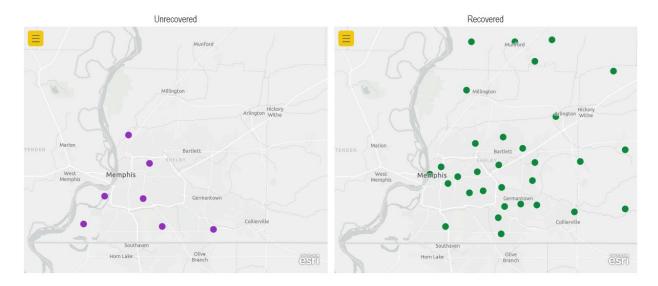


Figure 10. Maps of Memphis, TN, showing recovered (green) and unrecovered (purple) zip codes

Compare recovery status and trends for various price segments: The 90<sup>th</sup> percentile, median, and 10<sup>th</sup> percentile recovery status and trend over time could be analyzed and compared to each other to develop a more nuanced picture of market status at various price segments in each State.

Overlay socioeconomic data to this project's spatial analyses: The spatial data generated in this project could be compared to socioeconomic data from the Census Bureau to reveal correlations between recovery and socioeconomic status that are beyond the scope of this project or, conversely, it may reveal that these isolated unrecovered zip codes are in the decline stage of a normal real estate market cycle.

Analyze trends in specific zip codes: Additional trend analyses in unrecovered areas such as zip code 89108 in Las Vegas, NV, may prove useful, in that the zip code has not regained all of its pre-crash median owner occupied SFRD price, but trend analyses of the numbers of transactions shows the zip code to be a more active market than the pre-crash period, and the median sale prices are trending upward (Figure 11). A more finite trend analysis would allow a prediction of when a recovering zip code will return to its pre-crash levels. Similarly, an unrecovered zip code that has remained relatively flat for three or more years may indicate that the market in that area has established a new equilibrium, and a return to pre-crash levels may never occur.

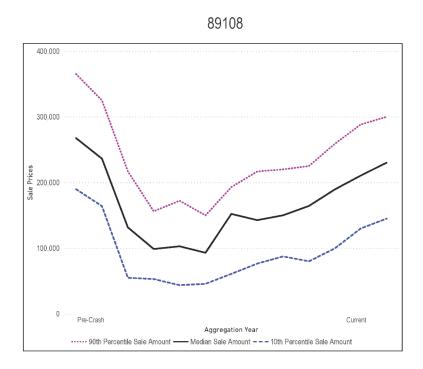


Figure 11. 90th percentile, median, and 10th percentile sale prices for each aggregation year for the 89108 zip code in Las Vegas, NV

FHWA may wish to consider additional analyses at the sub-State level to inform project-by-project Waiver determinations. Such analyses would be supportable assuming that the following issues are considered and criteria are developed to produce reliable conclusions under the given context.

Level of analysis: Analyses could be carried at varying levels, such as County, Metropolitan Statistical Area (MSA) and Combined Statistical Area (CSA), zip code, or some other sub-State geographic area. These could be used for project-by-project determinations, though it would need to be confirmed in each case that sufficient data are available.

*Data sufficiency:* Any analysis needs to be based on a defensible dataset. A zip-code level analysis needs to have adequate transactions in both the pre-crash and current periods to allow for comparison. The same is true for other levels of analysis. Some counties may have thousands of transactions, but others may have only a handful (and not enough to allow for defensible analysis).

*Varying recovery percentages:* A determination would need to be made about whether to treat two unrecovered areas the same way if they have significantly different recovery percentages. For example, should a zip code with a recovery percentage of 65% be treated the same as a zip code with a recovery percentage of 98%? Both are unrecovered, but one is much closer to full recovery.

Varying recovery status at different levels: A determination would need to be made about how to treat a given location if the recovery status varies across different levels of analysis. For example, how to treat a

given location that has not recovered at the zip code level, but has recovered when analyzed at the County and MSA level.

Geographic location of a project in a given zip code, MSA, or County. How should a project be treated if it is on the edge of an unrecovered zip code, and right up against several other zip codes that have recovered? Would a project on the other side of that zip code, and adjacent to unrecovered zip codes, be treated differently?

In summary, the data can generally support decisions at the sub-State level, though the smaller the geographic area, the more care must be taken to assure sufficient numbers of transactions. In general, as the geographic area increases (e.g., moving up from a zip code level to an MSA level to a State level), there will be a greater number of transactions and bigger sample size, resulting in a more defensible dataset for a given determination.

# Appendix: State Dashboards

