



# Understanding Data Anomalies in Consumer Behavior Data

A guide into Azira's approach in detecting and removing complex data anomalies in Consumer Behavior Data

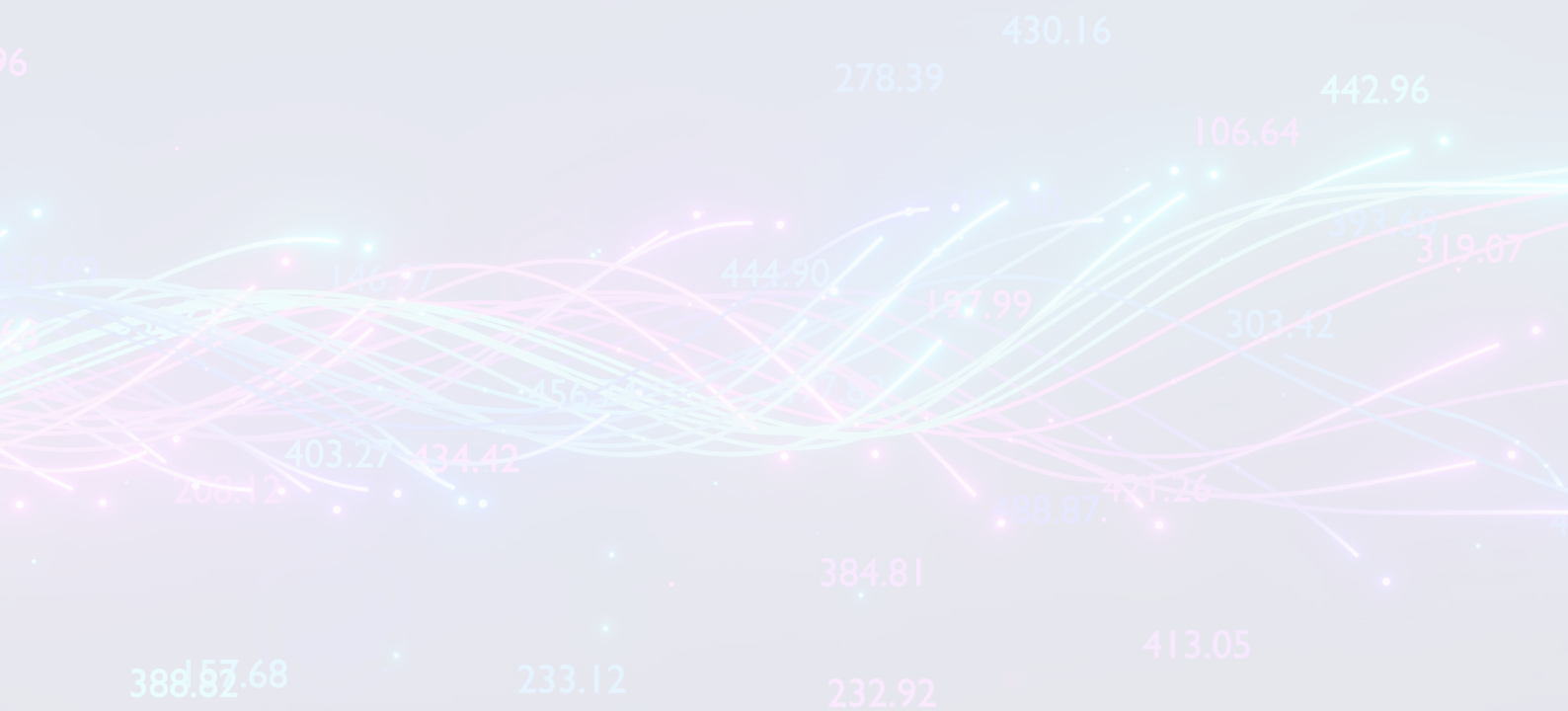
<b>Overview</b>	<b>1</b>
<b>Azira's Data Anomaly Filters</b>	<b>2</b>
<b>Detection and Removal</b>	<b>3</b>
<b>Conclusion</b>	<b>3</b>





# Overview

Azira collects data from a variety of sources. While the majority of our data has shown to accurately reflect consumer behavior, the mobile ecosystem continues to expand, and with it an increasing amount of anomalous data that can distort or misrepresent behaviors in location data. This white paper will outline some of the common types of anomalies that have been observed, along with some of our methodologies for identification and removal.





## Azira's Data Anomaly Filters

Anomaly Type	Description
Pauli	Large numbers of observations with different device IDs have the exact same latitude, longitude, and timestamp.
Crosshairs	Observations resemble a crosshair or plus sign pattern, though they do not actually fall on a grid.
Grid	<p>There are various grid anomalies that are similar in behavior. Devices that contribute to these anomalies appear to be copied from real device histories. Observations do not always fall exactly on the grid, leading to the theory that data sources sometimes resell data with some amount of jitter.</p> <ul style="list-style-type: none"><li>→ <b>2 KM Grid:</b> Observations fall on a 2km-spaced grid in a spherical CRS.</li><li>→ <b>Pow15 Grid:</b> Observations fall on a 0.0032768 decimal degree-spaced grid.</li><li>→ <b>Fiji Grid:</b> Observations fall on a grid of roughly 0.1 decimal degree spacing, originally found in Fiji data but appearing worldwide. It shows as a dashed line pattern.</li></ul>
Hot Circle	Observations from real devices and historical data fill circular areas 2km in diameter. These anomalies are similar to grid anomalies, as the data points for this anomaly often fall on a 0.0001-spaced grid (within a circular perimeter), but the circles themselves appear in random locations. This anomaly has been found in the US, Tokyo, Australia, and Russia.
Replay	A sequence of observations are "replayed" at a later time, sometimes resulting in impossible traffic patterns, such as store activity at a closed time/location, or visitation at a natural disaster site.



## Detection and Removal

Many of the anomalies listed above are due to duplicated data. The Pauli anomaly is essentially due to identical data points, and the grid anomalies are a result of duplicated data offset in a way that results in geometric patterns. We have developed algorithms for removing the anomalous data when the pattern is clear and relatively consistent; for example, we search for points along specific grids for grid anomalies, or points that form a large cluster for hot circles.

We are continuing to research and develop advanced methodologies for both detection and removal, incorporating image recognition techniques and statistical analyses, to ensure that our data quality remains consistent. However, as human movement, mobile device usage, and the mobile data ecosystem continue to evolve, we face a variety of challenges:

- New anomalies and fraudulent data appear in our sources on a regular basis. While we have implemented extensive monitoring, it can be difficult to pinpoint every new anomaly and develop additional filtering mechanisms.
- Aggressively removing anomalous data can impact data volume. For many of the anomalies we currently filter against, we look at the distribution of observations that are potentially anomalous, and remove groupings of points that are statistically unlikely. This balanced approach attempts to reduce the amount of discarded data, but the actual results can vary on a number of factors.
- Identifying anomalies at a large scale, particularly as they become increasingly complex, can be computationally expensive and time-consuming. For example, determining cases of the Replay Anomaly requires comparisons against patterns in historical data, which is non-trivial to develop and maintain.

## Conclusion

Data anomalies are a persistent problem in mobile data. At Azira, we are committed to high standards for data quality, and will continue to invest in the most current methodologies for protecting the value of our data.



Azira LLC, a global Consumer Insights platform, helps marketing and operational leaders improve their effectiveness with actionable intelligence to drive business results. Its mission is to create a more relevant world where brands are empowered to reach and build relationships with their consumers. With a profound commitment to partnership, trust and transparency, combined with decades of expertise in consumer behavioral analytics, Azira delivers innovative marketing solutions to curate audiences, activate omnichannel campaigns, and understand footfall attribution. It also provides operational insights for use cases such as site selection, trade area analysis, competitive intelligence and more. Azira serves enterprises in retail, hospitality, travel, real estate, financial services and media. A global company, Azira is headquartered in Los Angeles with offices in Paris, Bangalore, Singapore, Sydney, and Tokyo. To learn more, please visit <https://azira.com>.