

ViziSense.com

Audience Measurement Methodology

Table of Contents

Introduction	3
Population Definitions	4
Audience:	4
Home:	4
Office:	4
Cybercafé:	4
Schools / Colleges:	4
Data Collection	5
Panel Recruitment	5
Panel Verification.....	5
Panel Data Collection	6
Pixel Data Collection.....	6
Data Security and Anonymity	7
Monthly Sample Selection	8
Pixel Data Selection.....	8
Data Analysis – Weighting and Projections	9
Overall Population Estimates	9
Overall Population Distribution	9
Bias Elimination.....	11
Demographics.....	11
Internet Usage:	11
Confidence Measure	12
Reporting	13
Reach and Frequency Methodology	14

Introduction

ViziSense (www.vizisense.com) is India's first free and open online audience measurement platform that provides advertisers, agencies and publishers with demographic and traffic data for any website.

ViziSense is an exhaustive audience measurement platform that can be used, among other things, to look up audience and traffic estimates for any internet site accessed by Indian audience. Advertisers and media agencies can benefit with improved audience targeting and the ability to plan more accurate media buys. Publishers can provide more accurate data about the audience engaging with their content, allowing advertisers to spend more with them. ViziSense also acts as a research tool for competitive benchmarking and consumer behavior analysis.

ViziSense sources site-specific data from

1. a well diversified panel of 55,000 active Indian internet users and
2. ViziSense tracker tag (Pixel) that is embedded by publishers, thus providing actual traffic data for a given website.

ViziSense then employs various statistical techniques to triangulate and scale this data into accurate population level estimates for websites.

Population Definitions

Audience:

Internet users residing in India accessing any website in the world in the last 30 days.

Home:

People accessing internet from home, excluding those who are running a business operated from their residence, are classified as Home users.

Office:

Those accessing internet from their office computer, including a business operated from home, are considered Office users.

Cybercafé:

A user is classified as a cybercafé user if she is accessing the internet from a cybercafé. This excludes users who use a personal computer at home or office. There is no intersection between a cybercafé user and a home/office user.

Schools / Colleges:

A user is classified into this group if she is accessing the internet from places such as schools, colleges or universities.

Data Collection

ViziSense runs a well diversified panel of 55,000 active opt-in Indian internet users who allow ViziSense to anonymously and securely capture their web surfing behavior. These members have given ViziSense explicit permission to confidentially monitor their online activities in return for attractive incentives including gifts and prizes.

Another parallel source of site specific data is ViziSense tracker tags (also termed Pixel in this note) that are embedded by publishers into web pages allowing ViziSense to record every request to their sites. The tags are used not only to measure census data such as page views, visits, unique users, reach and geography but also to predict demographic and behavioral characteristics of the audience visiting the website. No Personally Identifiable Information (PII) is captured during the entire process of tracking browsing activity via ViziSense tracker tags.

Panel Recruitment

Low penetration of Internet in India renders Random Digit Dialing an unviable process to recruit a large panel. Hence, ViziSense has focused on online acquisition techniques such as direct mailers, online ads and promotions for recruiting the panel. Online acquisition is carried over a wide spectrum of websites appealing to a diverse cross section of India's online population, in order to ensure a well diversified and representative panel. To minimize self-selection bias, appropriate incentives are offered for each segment. To address potential self-selection bias resulting from online sign-up, ViziSense further calibrates the panel using *overall* population distribution judiciously derived from various independent reports. (See section entitled "Data Analysis – Weighting and Projections")

ViziSense considered two approaches to calibrate the panel:

1. Employ weighting techniques for computing projection weights for each panel member or panel segment.
2. Optimize existing panel size and composition to reflect population distribution by deselecting 'extra' panel members from the active sample used for computation.

ViziSense selected the first approach as the latter would lead to significant information loss.

Panel Verification

Users who choose to be part of the ViziSense panel are asked a series of questions about their household, demographics, disposable income and purchase intent. ViziSense periodically calls a random subset of the panel to verify these details and to update dynamic information. As a verification process, panel members are sent a notification intermittently to resubmit demographic details in order to cross-check information submitted previously. This is used to gray-list panel members who may have provided incorrect information. However, their browsing behavior is used for traffic projections, but not for estimating demographics. The

verification process also prevents dynamic demographic information such as income from getting outdated.

Over the course of many verification cycles, it has been found that a vast majority of panel members have provided accurate information.

Panel Data Collection

In order to join the ViziSense panel, users download and install an application that passively captures internet activity of the user and relays it batch-wise to ViziSense servers. This application currently captures data for IE5+ and FireFox1.5+ browsers running on Win32 Operating System. This activity is captured even if a user is behind a firewall and is independent of connection speed and ISP. The application captures only the URLs visited, and does not capture or relay sensitive information such as products purchased, credit card details and online banking account information. The panel data transfer from user's computer to the ViziSense server has negligible impact on the user's bandwidth consumption.

The browsing activity of the user is captured in a local file on the users' computer, and the file is regularly sent to ViziSense servers. The files sent to the server are appropriately encrypted, as is the data stored on ViziSense's access restricted servers.

ViziSense does not capture any other internet activity on a users' computer including torrent downloads, IM applications or other digital applications on the desktop. Currently, ViziSense does not capture Ajax requests or 3rd party widgets distributed on the web, but the measurement of these would be supported in upcoming versions soon.

Pixel Data Collection

ViziSense tracker tag (pixel) acts as a parallel source for capturing site traffic and triangulating final estimates. Participating publishers are provided with a ViziSense tag which can be placed on their web pages. A request is made to the ViziSense server whenever a page containing ViziSense tag is loaded on a user's browser. This request includes information such as browser used and URL visited, and is anchored against 3rd party browser cookie set by ViziSense.com. ViziSense servers that receive tracker tag request maintain high performance and availability, and may use a Content Delivery Network (CDN) service.

Data Security and Anonymity

The anonymity of all personnel in any way concerned with sample respondents or households is preserved and protected.

The panel application captures only the URLs visited, and does not capture or relay sensitive information such as products purchased, credit card details and online banking account information.

No Personally Identifiable Information (PII) is captured during the entire process of tracking browsing activity via ViziSense tracker tags.

The computer programs and data are diligently protected from unauthorized manipulation.

Monthly Sample Selection

For panel based estimates, ViziSense does not consider its entire panel while calculating various statistics. Not all people recruited on the panel are active each month. Also, there is expected monthly attrition as some of the existing members leave, and the panel is replenished with new ones.

The active panel constitutes those who have accessed the internet at least one day during the entire month. A subset of these are selected for computing a given month's estimate based on the following –

1. Panelist belongs to the country (India) for which data is being calculated for.
2. Panelist has used the internet at least once during the month.
3. First reported activity for the panelist was recorded at least thirty days ago. This is to ensure that new panel members joined in the month of reporting do not skew panel estimates.
4. Panelist, if classified as home or office, uses a computer for which major portion of the activity can be attributed to a single user.
5. Panelist hasn't left the panel after recording activity in the initial part of the month – statistically estimated based on the average inactive period of the user (defined as the duration between any consecutive active days) based on the past two month's trend.

For demographic estimates, only those members of the subset are considered who are most active from home or office computer.

ViziSense closely monitors panel activity for unusual patterns to identify and eliminate any attempts to bias measurement results by a particular respondent's submission of fabricated information.

Pixel Data Selection

ViziSense has built robust fraud detection techniques, and constantly works on enhancing them to address potential 'click-fraud' by certain ill-advised publishers wanting to artificially inflate site stats. Sudden, disproportionate spike in browsing activity on a particular website from one or a few limited IP addresses generates an alert that is manually verified to see if the spike is a generic trend or not. These stats are also continually compared with stats from the ViziSense panel to sniff for unusual patterns.

Site estimates from the ViziSense Pixel source are adjusted with expected cookie deletion rates. Prior independent studies including those conducted by Jupiter Research in 2005 and by Nielsen/Net Ratings in 2005 conclude that cookies are deleted by at least 30% of Internet users in a month. ViziSense is working on building more advanced and locally relevant models to account for cookie deletion in the Indian context.

Data Analysis – Weighting and Projections

To address potential self-selection bias due to online panel recruitment, ViziSense calibrates the panel using overall population distribution judiciously derived from various annually published 3rd party surveys and reports. Since these population distributions come from different sources, ViziSense obtains projection weights through a proprietary Iterative Proportional Fit algorithm. The feedback loop incorporated herein finds the best fit value for *overall* population distribution derived from panel-based relative distribution for universally appealing portals, various independent reports, and further validates it by running Monte Carlo simulations on randomly created test sub-samples for iterative improvements.

Based on the above methods, ViziSense estimates for India’s online population are as follows.

Overall Population Estimates

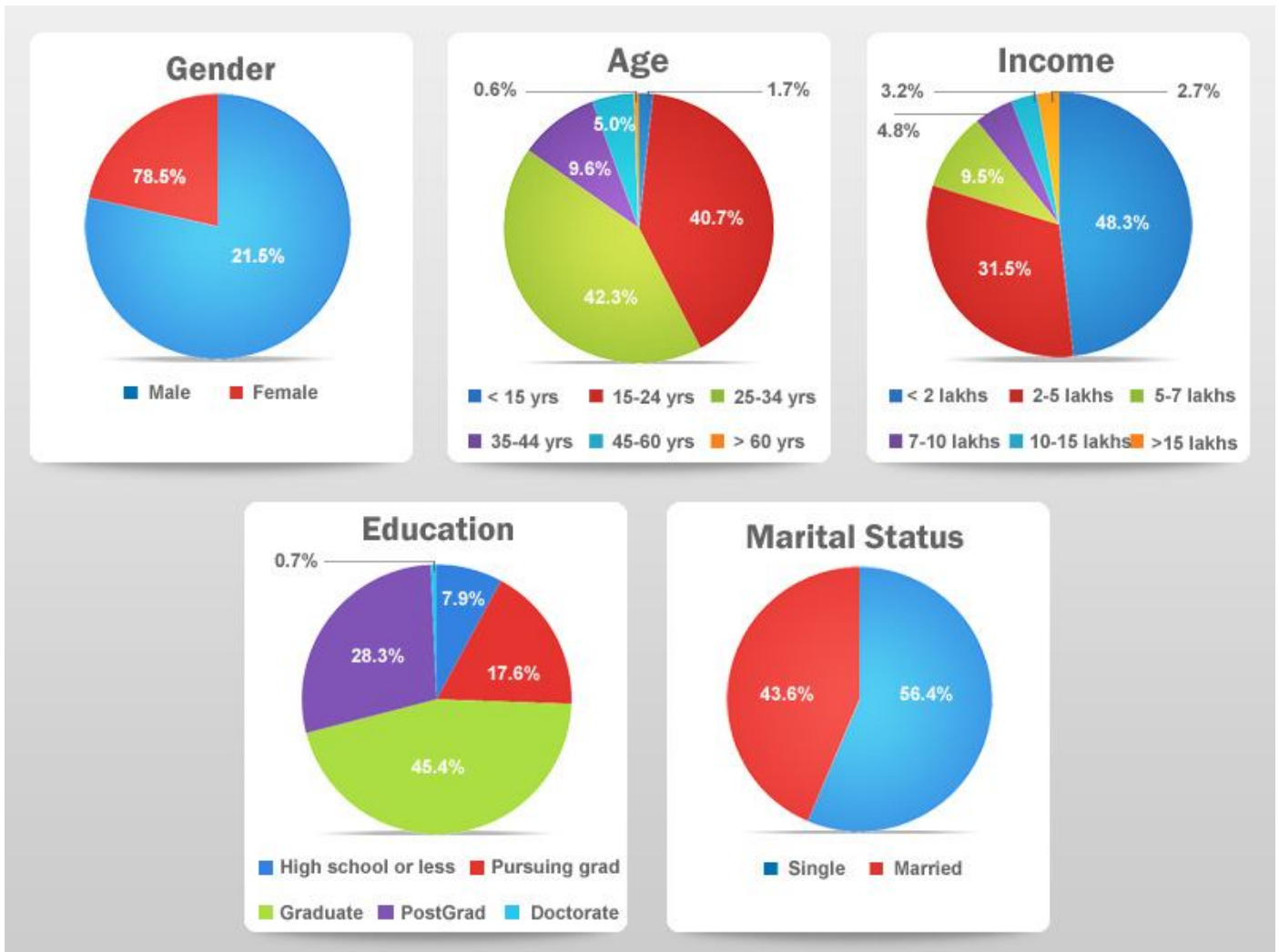
Active online population (Sep 2009)	40 million
Home users	12 million
Office users	9 million
Home and Office users	3 million
Cybercafé users	13 million
School / college users	3 million

Overall Population Distribution

The following table contains composition of the selected panel for Sep 2009, and how it compares with estimated population composition (or distribution). Accordingly, projection weights are assigned to the active panel to estimate population stats for a given site by scaling panel statistics.

Category	Population Distribution	Selected Panel (Sep 2009)
Gender		
Male	80.2%	78.5%
Female	19.8%	21.5%
Age		
Less than 15 yrs	0.8%	1.7%
15-24 yrs	40.4%	40.7%
25-35 yrs	31.6%	42.3%
36-44 yrs	15.3%	9.6%
45-60 yrs	9.2%	5.0%
Above 60 yrs	2.7%	0.6%
Education		

High School or Less	12.1%	7.9%
Pursuing Graduation	22.5%	17.6%
Graduate / Bachelors	38.6%	45.4%
Post Graduate / Masters	24.3%	28.3%
PhD / Doctorate	2.5%	0.7%
Household Income		
Less than Rs. 2 lakh	55.8%	48.3%
Rs. 2 lakh – Rs. 4.99 lakh	22.9%	31.5%
Rs. 5 lakh – Rs. 6.99 lakh	9.3%	9.5%
Rs. 7 lakh – Rs. 9.99 lakh	5.1%	4.8%
Rs. 10 lakh – Rs. 14.99 lakh	2.9%	3.2%
Greater than Rs. 15 lakh	4%	2.7%
Marital Status		
Single	51.2%	56.4%
Married	48.8%	43.6%



Graphical Representation of Panel Composition

Bias Elimination

Panel users are categorized into appropriate clusters and assigned weights using estimated population distribution listed above. These clusters are based on following parameters.

Demographics

1. Gender
2. Age
3. Education
4. Household Income
5. Marital Status

Internet Usage:

1. Usage segments (Light, Moderate and Heavy) based on average monthly page views by a given user across all sites, measured in the last two months of activity.
2. Interest segments based on primary category of sites visited in the last two months (e.g. tech, games, etc.).

Confidence Measure

ViziSense assigns a confidence measure to overall statistical estimates for every site. This depends on the number of the selected panelists who visit the site in a given month. Assuming Gaussian distribution, confidence measure is approximated by the number of users who must visit a website for a certain level of confidence. Further enhancements to the confidence measure model are being implemented.

If a site gets less than 10 unique users from the selected panel, no data is reported back for the website. This is done to eliminate sites for which the number of visitors is so low that even a minor fluctuation in these numbers could lead to a dramatic change in the overall statistics.

For sites that have embedded a ViziSense Pixel in all their pages, the confidence measure is very high for overall traffic, as the Pixel captures census level data for that site.

Reporting

ViziSense.com publishes the following reports, updated every calendar month, across tens of thousands of sites (and their sub-domains) accessed by the Indian audience.

1. Audience Distribution
 - a. Distribution by Gender, Age, Income, Education
 - b. Composition Index – i.e. how a site’s audience composition compares with overall India composition on a particular parameter.
2. Site Visits Summary
 - a. India rank
 - b. Unique visitors and % reach
 - c. Monthly visits and page views
 - d. Average monthly visits per visitor
 - e. Average page views per visitor
 - f. Average time spent per visit (in seconds)
 - g. Daily traffic trend for the month – visits and page views
3. Traffic Distribution
 - a. Distribution by top cities & states
 - b. Distribution based on point of access – home, office, cybercafé, schools / colleges
 - c. India vs. non-India traffic data – only for publishers carrying ViziSense tracker tag
4. Data across all Sub-domains – unique visitors, reach, visits, page views per month
5. Top 100 domains report – unique visitors, reach, visits, page views per month
6. MediaMix – The media planning product from ViziSense identifies the right set of sites for a given audience query across age, gender, income, education and location.

For all its reporting, ViziSense undertakes diligent editorial efforts, to the best extent possible, to block offensive sites including porn, terror, hatred and abusive sites from getting measured and ranked.

Reach and Frequency Methodology

The reach and frequency planning tool allows the planners to identify the unique visitors, reach and frequency their campaign would achieve over a specified duration and target impressions.

The inputs taken from the planner are as follows:

1. Target Audience on parameters of Age, Gender, Income, Education and Location
2. Target set of sites for the campaign
3. Impressions that he wishes to buy on each of the sites. It is assumed that he wishes to target only one ad impression per page.
4. Whether the site allows Audience targeting or not
5. Expected CPM for each site for the given impressions and target audience.

The following parameters are available as an output:

1. Unduplicated unique visitors that shall be exposed to the campaign, on each site as well as cumulative.
2. The number of unique visitors as a percentage of the total web population for the selected demographic filters
3. Frequency i.e. the average number of times the users shall be exposed to the campaign during the entire duration of the campaign.
4. Total cost on each of the sites, which is a factor of the impressions bought and the CPM pricing for each site.
5. Cost per 1000 users, calculated by using the total cost and the expected unique visitors on each site

ViziSense shows the above output parameters for each of the sites selected by the planner. In addition, the output for all the sites combined together is also displayed in another row as a Summary of the entire plan. The Summary output incorporates the duplication among sites considered, and thus, shows only the cumulative unduplicated unique reach the entire campaign would deliver.

The following factors are factored in the computation of the output:

Page view consumption pattern:

The main variables governing the computation of the expected unique visitors are the impressions bought and the duration of the campaign. In order to account for the page view consumption patterns and the activity of the users by every passing day, the panel is studied to draw trends of the increase in unique visitors with increase in the impressions bought and increase in the duration of the campaign. However, if one were to buy all available impressions on a site, he will reach the entire audience that visits the site for the specified duration.

Effective Page Views:

As stated above, the reach and frequency is calculated for a selected set of audience that the planner selects. Therefore, it becomes important to identify the exact number of impressions that can be associated to the target group. Since there will be spill over of the impressions to the other users as well, the estimated impressions available for the TG is calculated by factoring the ratio of the page view consumption of the TG and the total audience and checking if the site allows audience targeting or not.

Overlap in Time:

For a given campaign duration, with each new day, there are new users that come on to a site. This means the impressions bought will be exposed to newer users by each progressing day. However, even those users who visited a particular site earlier but were not exposed to the campaign are still eligible to be exposed for the 1st time and are therefore treated as new users. The unique visitors that are exposed to the ad are calculated by considering this pool of new and old unexposed users.

Overlap between sites:

The Summary row shows the unduplicated unique visitors that the overall campaign would achieve for the planner. Since a user could be common to two or more sites, the contribution of each site to the common pool of users is used to calculate overlap among users who may have seen the ad on more than one site. The overlap typically, increases with the increase in the number of impressions bought for the same campaign duration.

Variable Campaign Durations:

A planner has the flexibility to specify the duration of the campaign of each of his selected sites. These durations could be different for different sites and thus the unique visitors that shall be reached is computed after accounting for the page view consumption patterns over the specified durations for each of the sites. It is assumed that the campaign starts on the same day on each of the sites selected and runs as per the specified duration for that site. After the campaign on a particular site is over, the other sites on which the campaign is live will keep adding to pool of users who are exposed to the campaign. The total unduplicated unique visitors for the entire campaign is calculated by identifying the number of new unexposed set of users that each active site incrementally adds after the campaigns on the other sites have ended.