جامعة نيويورك أبوظبي
NYU | ABU DHABI

16 May 2023

A Comparative Evaluation of Next-Generation Image Formats on Low-Cost Mobile Hardware

Advisor

Professor Yasir Zaki

Student

Shaurya Singh

Capstone Defense

جامعة نيويورك أبوظبي
NYU | ABU DHABI

Table of Contents

I Research Background & Motivation

II Related Work

III Methodology

IV Results

Introduction

- The number of internet users is exploding. Google's Next Billion Users research project estimates that within the next four years, a billion more people will purchase their first smartphone.

- HTTP Archive's set of over 1 million web pages using BigQuery suggests that 82% of web-pages use at least one PNG image, whereas a whopping 91% of web-pages in the set use at least one JPEG image.

- The median mobile page weight has increased from 295KB in January 2012 to 1993KB in April 2023, an increase of 575% in slightly over a decade.

- The 75th percentile mobile page weight stands at 4062KB, and this number rises to 8195KB for the 90th percentile. The median mobile web page contains 677KB of image data, which accounts for over a third of all page weight,

The WebP Image Format

- The WebP (Web Picture) image format, developed by Google, represents an open, royalty-free solution designed to replace existing image formats like JPEG, PNG, and GIF.
- WebP utilizes the modern VP8 image compression algorithm, enabling it to deliver high-quality images with significantly reduced file sizes.
- The WebP format supports a wide range of color depths and alpha transparency. It also supports animation, making it a potential replacement for the ageing GIF standard in the future.
- WebP is supported by many desktop and mobile browsers, including popular ones like Chrome, Edge, and Mozilla.

The AVIF Image Format

- The AVIF (AV1 Image File Format) is an new image file format that utilizes the advanced AV1 video codec to compress and encode image data. This cutting-edge format was developed by the Alliance for Open Media (AOMedia), with the vision of providing an open, royalty-free alternative to existing, often restrictive, image formats such as JPEG, PNG, and GIF.
- AVIF also supports a broad range of color depths and is compatible with alpha transparency and allows for animation.
- Extremely efficient image encoding, but can be processing power-hungry when decoding.

Previous work on WebP

- WebP has a significantly better data compression algorithm compared to JPEG and PNG, for only slightly worse compression quality. (Si and Shen, 2016)

- WebP shows notable efficiency when it comes to compressing natural images, often slightly outperforming other image codecs. This could make it an effective choice for applications where natural images need to be compressed to lower bitrates. (Ginesu et al, 2012)

- Facebook has slowly phased out JPG/PNG standards on both their desktop and mobile platforms, instead using WebP.

Previous work on AVIF

- AVIF “results in the best overall performance [and bitrate savings] considering both 4:2:0 and 4:4:4 chroma sub-sampling encoded images” and also outperforms WebP. (Barman et al, 2020)

- n 2020, Netflix began moving most of their image assets to an AVIF-first approach, with JPEG and PNG files as fallback options. Netflix’s large scale experimentation with AVIF concluded "superior compression efficiency" compared to JPEG, with similar picture quality in the end product.

Selection of Analyses

- We wish to study the comparative gains offered by next-generation image formats for webpages originating in high and low income countries. For this purpose, we devise a Country-wise Analysis detailed in the next sections.

- Furthermore, we wish to see how these formats largely affect the top webpages in the world, for which we devise a Representative List Analysis.

- Finally, we study the Image-Level impacts to isolate our study to images only, and not webpages.

Country-wise Analysis - Sampling URLs

For the purposes of this project, we have used the Tranco list for the top 1 million web pages in the world. This list is more temporally stable and uses data aggregated from the Alexa, Cisco Umbrella, and Majestic lists. Each of these lists has its own strengths and weaknesses which are briefly described below:

- Alexa ranks web pages based on page visits and is extremely volatile.
- Cisco Umbrella tracks DNS traffic and is also volatile, although less temporally volatile than Alexa. Since it uses DNS traffic, it also includes pages that are not visited through a browser (such as subdomains hosting cloud-based databases). In general, recording subdomains as distinct entries in the list is a big weakness for Cisco Umbrella.
- Majestic tracks backlinks to a domain and is therefore extremely temporally stable, even to a fault. Well-ranked websites will often retain their positions well after their popularity has faded due to functional backlinks.

Tranco combines these lists using Dowdall's rule, where the first domain gets 1 point and the nth domain is assigned 1/n points. By combining the scores of domains across constituent lists, Tranco generates a more reliable ranking for the world's most popular web pages.

Country-wise Analysis - Sampling URLs

We use the following configurations to generate a custom Tranco list:

- Each domain included in the custom list should be a part of at least 2 out of 3 base lists.
- We exclude subdomains from our analysis.
- For organizations that have many local TLDs (such as Google), we only consider the single most popular URL and discard the rest. This analysis is predominantly focused on comparing local homegrown web pages for each country/region, and local subdomains operated by large multinational corporations can warp the results of our analysis.
- We also restrict our domains to those that appear in the Chrome User Experience Report of 2023, which contains URLs popular with Chrome users and therefore reflects browser traffic across a large user base. It also allows us to standardize our own tests of image optimization techniques, which are also run on Chrome.

Setting these configurations returns a more robust list of popular web pages that is 366,410 entries long. Note that the initial Tranco list using the base parameters is over a million entries long.

Country-wise Analysis - Sampling Countries

For this project, we need to sample a diverse set of countries that vary from each other significantly in terms of economic benchmarks. We use the World Bank's GNI (US$, calculated using the Atlas method) metric to classify countries into 4 categories:

- Low-Income Economies
- Low-Middle Income Economies
- Upper-Middle Income Economies
- High-Income Economies

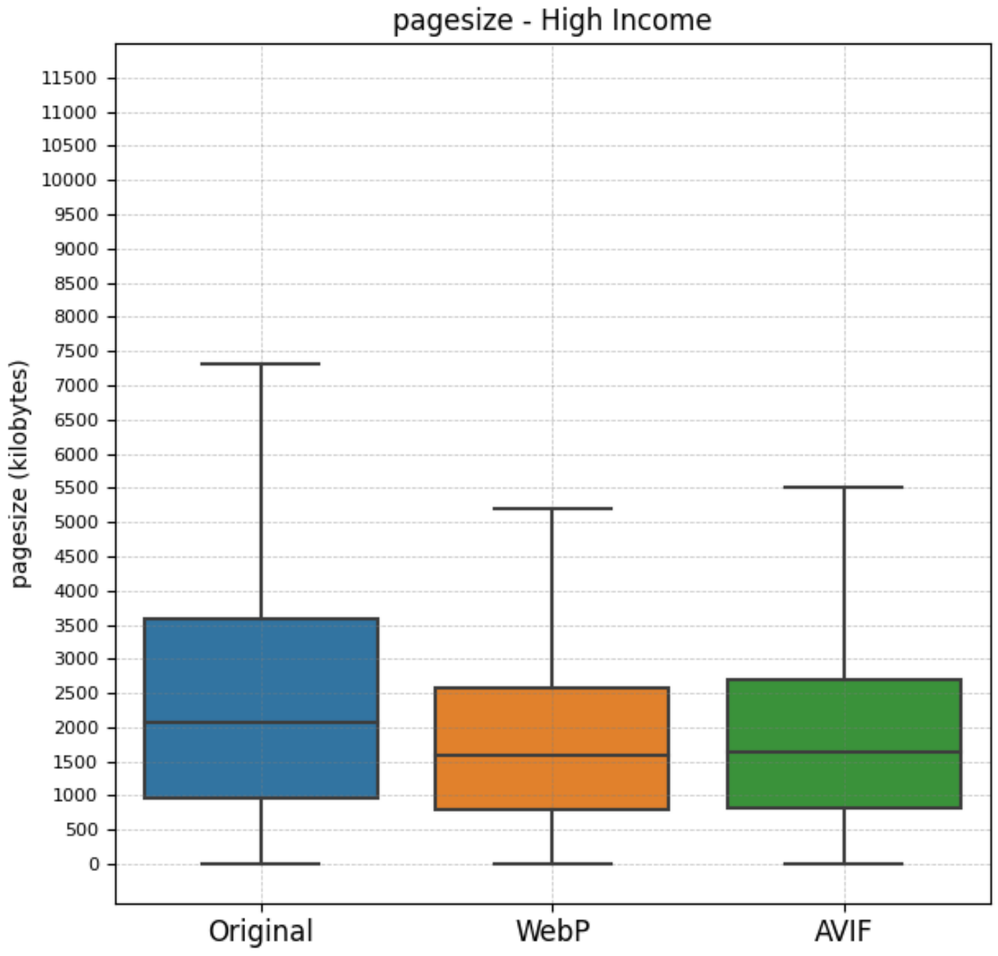
Furthermore, we control for population by picking the 30 most populous countries in the world.

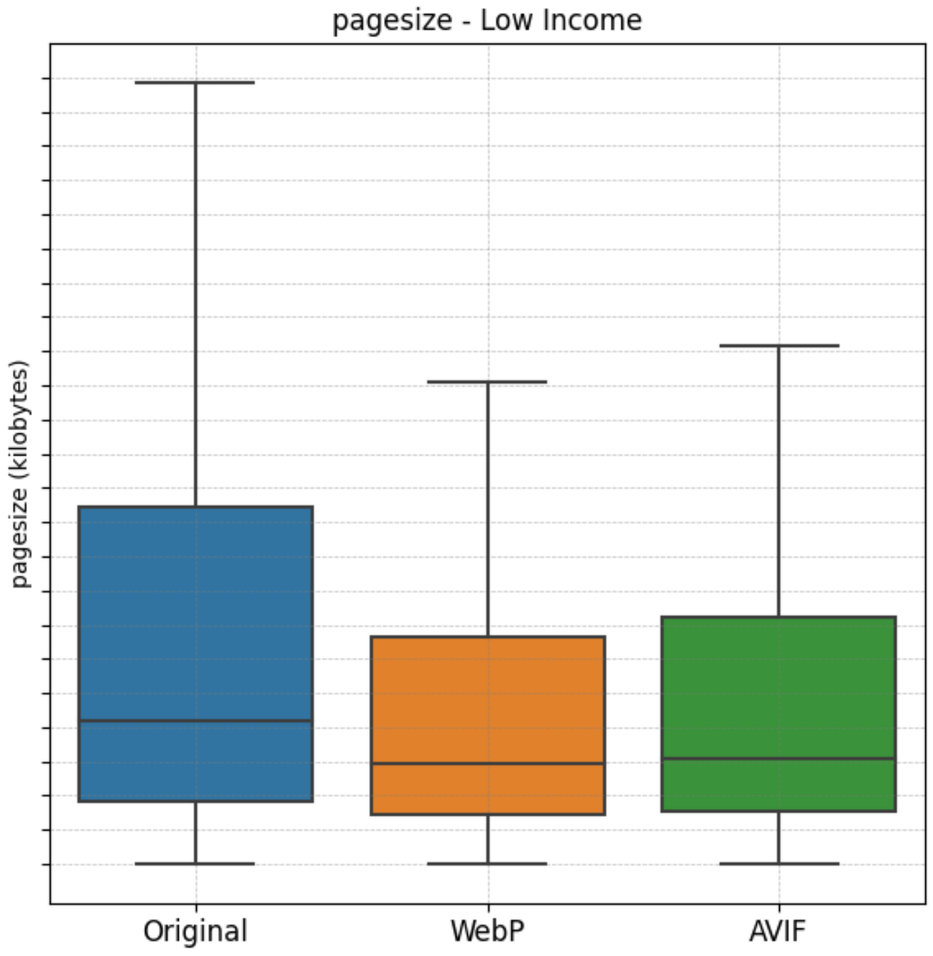
These countries give us a good mix of the above categories; in fact, the top 30 most populous countries in the world are an almost even 50/50 mix of low/low-middle and high/upper-middle joint buckets. The population threshold for a spot in the top 3 also neatly fits in at 50 million. Combined, these countries account for over 85% of the world's population, thereby giving our dataset increased relevance.

Country-wise Analysis - Sampling Countries

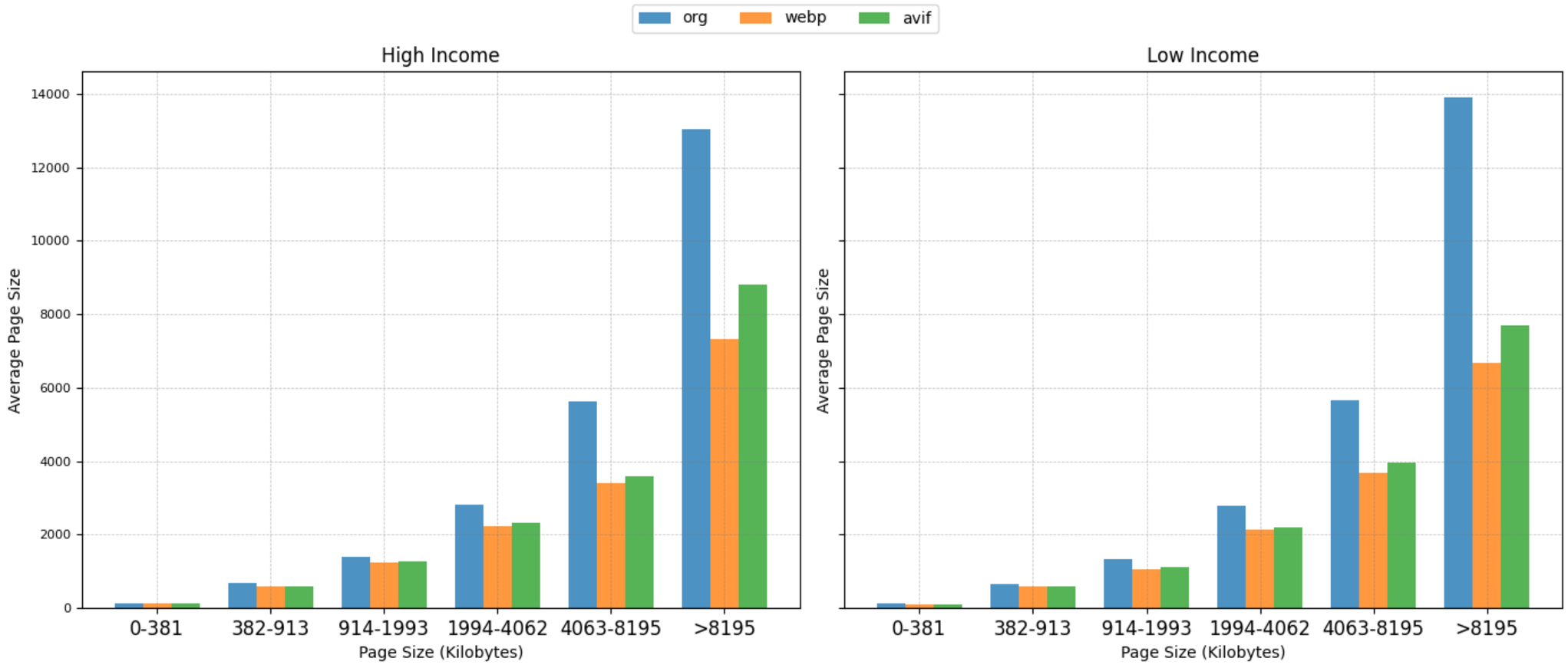
#	Country	Population	GNI per capita classification	Region	TLD	
1	Myanmar	54,045,420	a. Low Income Economy	East Asia a…	.mm	
2	Vietnam	96,462,106	a. Low Income Economy	East Asia a…	.vn	
3	Philippines	108,116,615	b. Lower-middle Income Eco…	East Asia a…	.ph	
4	Indonesia	270,203,917	b. Lower-middle Income Eco…	East Asia a…	.id	
5	Egypt	100,388,073	b. Lower-middle Income Eco…	MENA	.eg	
6	Iran	82,913,906	b. Lower-middle Income Eco…	MENA	.ir	
7	Pakistan	216,565,318	b. Lower-middle Income Eco…	South Asia	.pk	
8	Bangladesh	167,310,838	b. Lower-middle Income Eco…	South Asia	.bd	
9	India	1,366,417,754	b. Lower-middle Income Eco…	South Asia	.in	
10	Tanzania	58,005,463	b. Lower-middle Income Eco…	Sub-Sahar…	.tz	
11	Kenya	52,573,973	b. Lower-middle Income Eco…	Sub-Sahar…	.ke	
12	Nigeria	200,963,599	b. Lower-middle Income Eco…	Sub-Sahar…	.ng	
13	Democratic Republi	86,790,567	b. Lower-middle Income Eco…	Sub-Sahar…	.cd	
23	Thailand	69,625,582	b. Lower-middle Income Eco…	East Asia a…	.th	
14	Ethiopia	112,078,730	b. Lower-middle Income Eco…	Sub-Sahar…	.et	3,042,461,861
15	South Korea	51,709,098	d. High Income Economy	East Asia a…	.kr	
16	Japan	126,226,568	d. High Income Economy	East Asia a…	.jp	
17	Spain	47,076,781	d. High Income Economy	Europe an…	.es	
18	Italy	60,297,396	d. High Income Economy	Europe an…	.it	
19	France	67,059,887	d. High Income Economy	Europe an…	.fr	
20	United Kingdom	66,834,405	d. High Income Economy	Europe an…	.uk	2,903,565,761
21	Germany	83,132,799	d. High Income Economy	Europe an…	.de	
22	United States	328,239,523	d. High Income Economy	North Ame…	.us	
24	China	1,397,715,000	c. Upper-middle Income Eco…	East Asia a…	.cn	
25	Turkey	83,429,615	c. Upper-middle Income Eco…	Europe an…	.tr	
26	Russia	144,373,535	c. Upper-middle Income Eco…	Europe an…	.ru	
27	Colombia	50,339,443	c. Upper-middle Income Eco…	Latin Amer…	.co	
28	Brazil	212,559,417	c. Upper-middle Income Eco…	Latin Amer…	.br	
29	Mexico	126,014,024	c. Upper-middle Income Eco…	Latin Amer…	.mx	
30	South Africa	58,558,270	c. Upper-middle Income Eco…	Sub-Sahar…	.za	

Country-wise Analysis

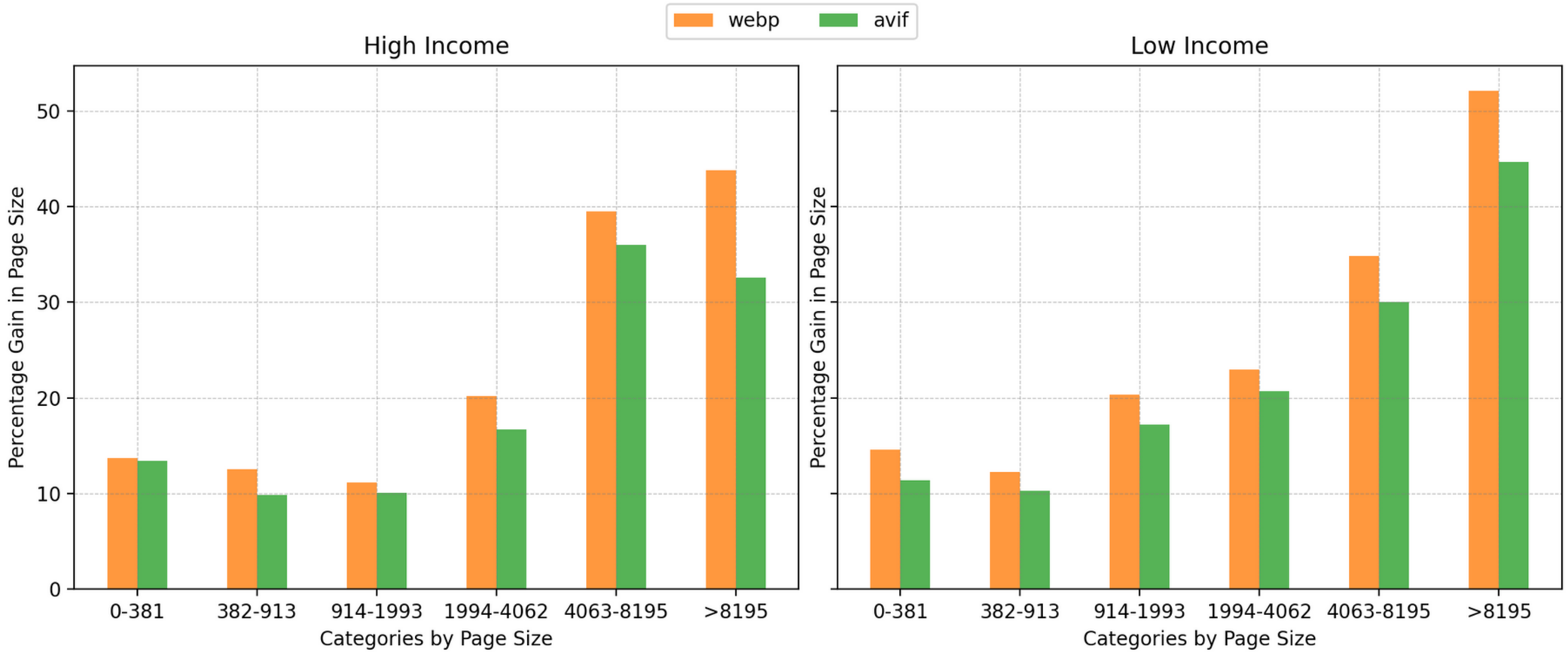




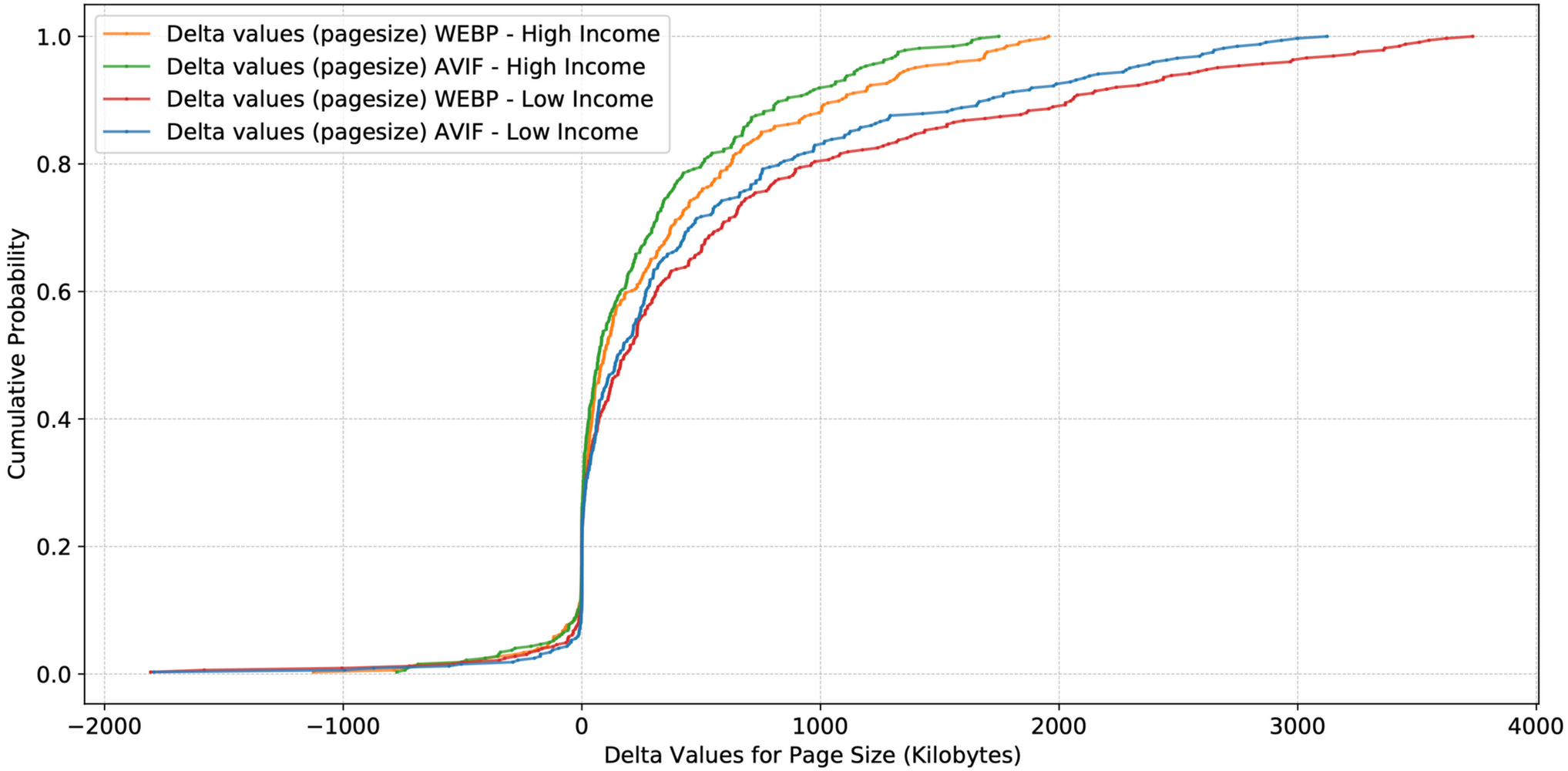
Country-wise Analysis: Data Analysis



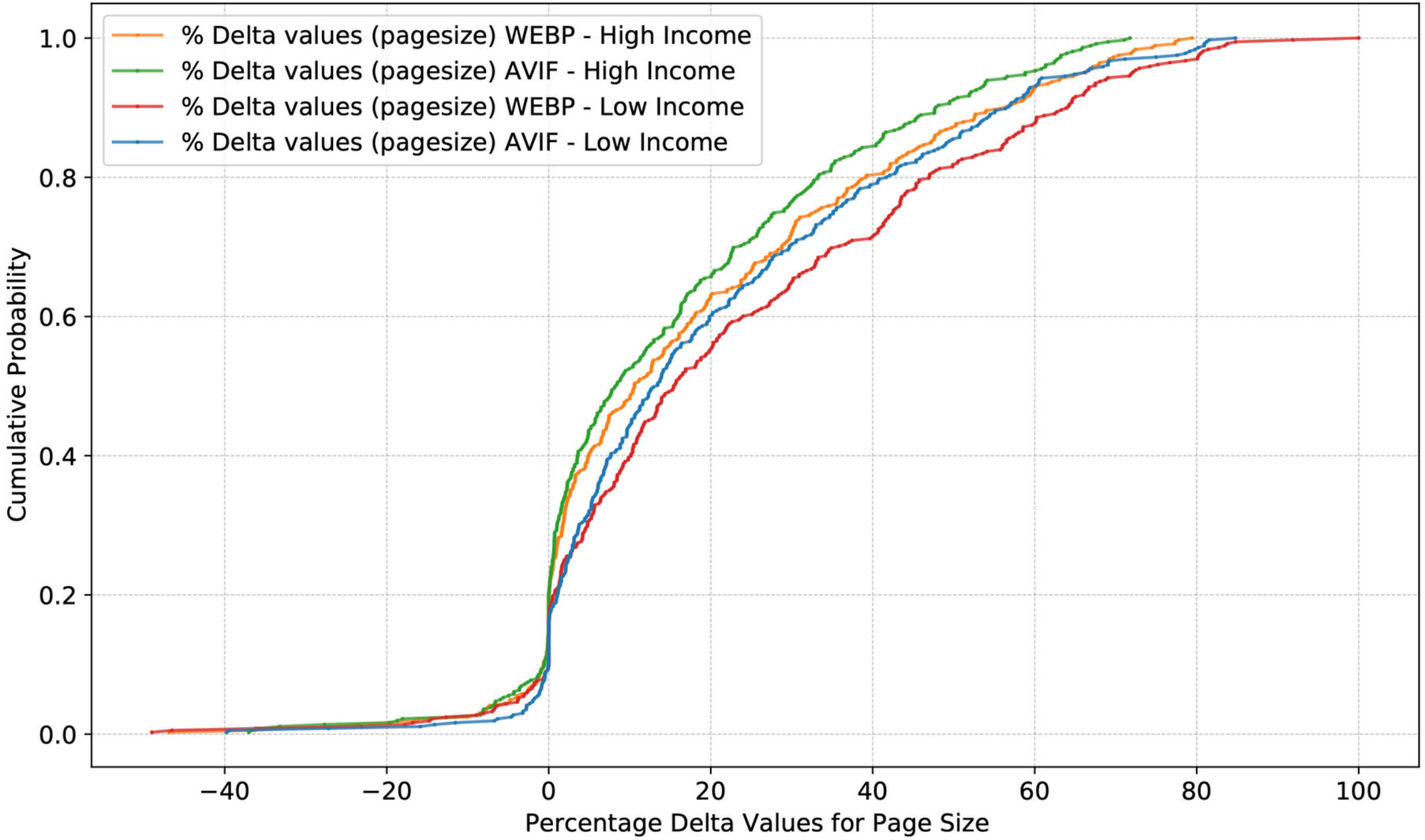
Country-wise Analysis: Data Analysis



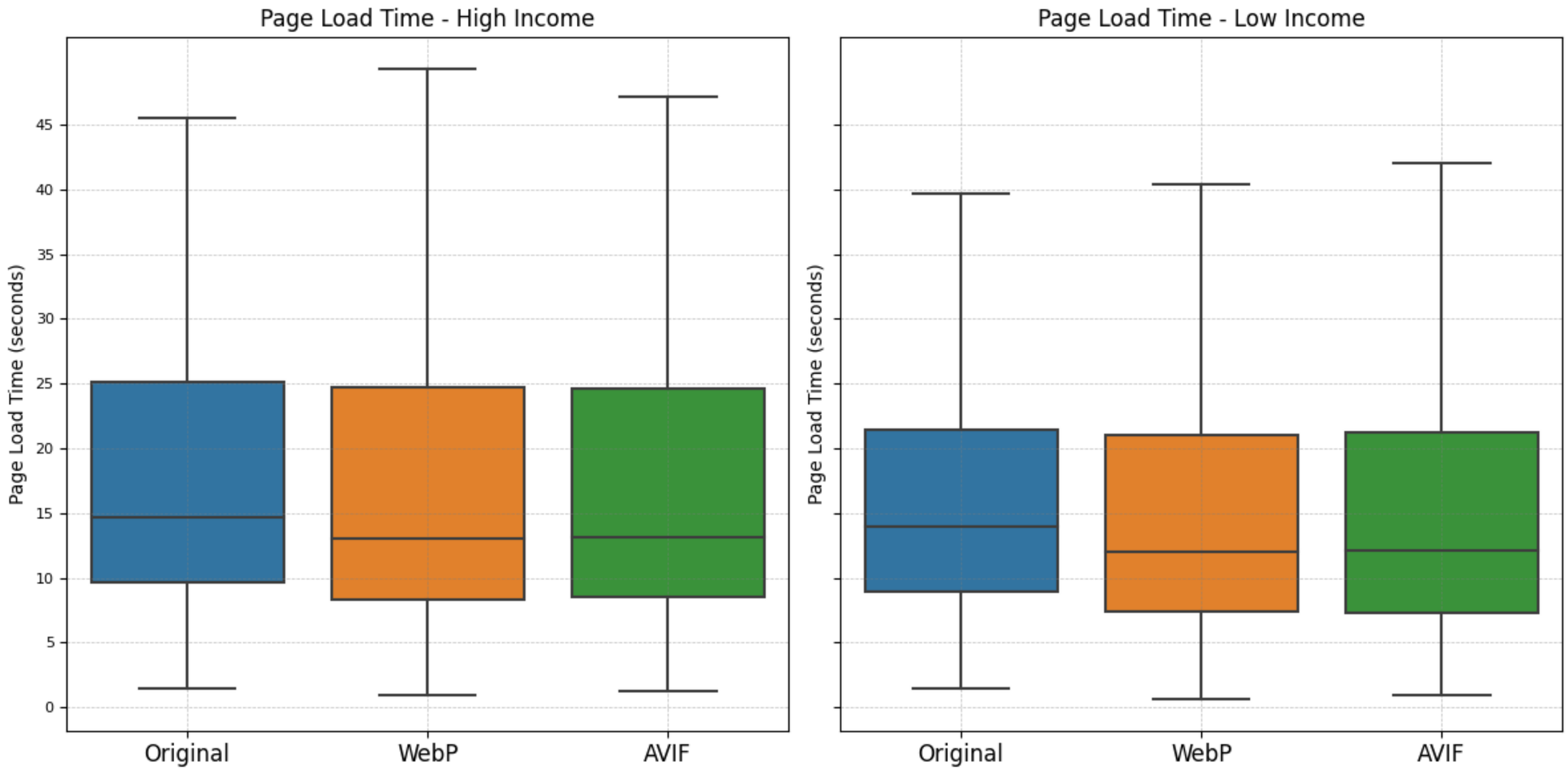
Country-wise Analysis



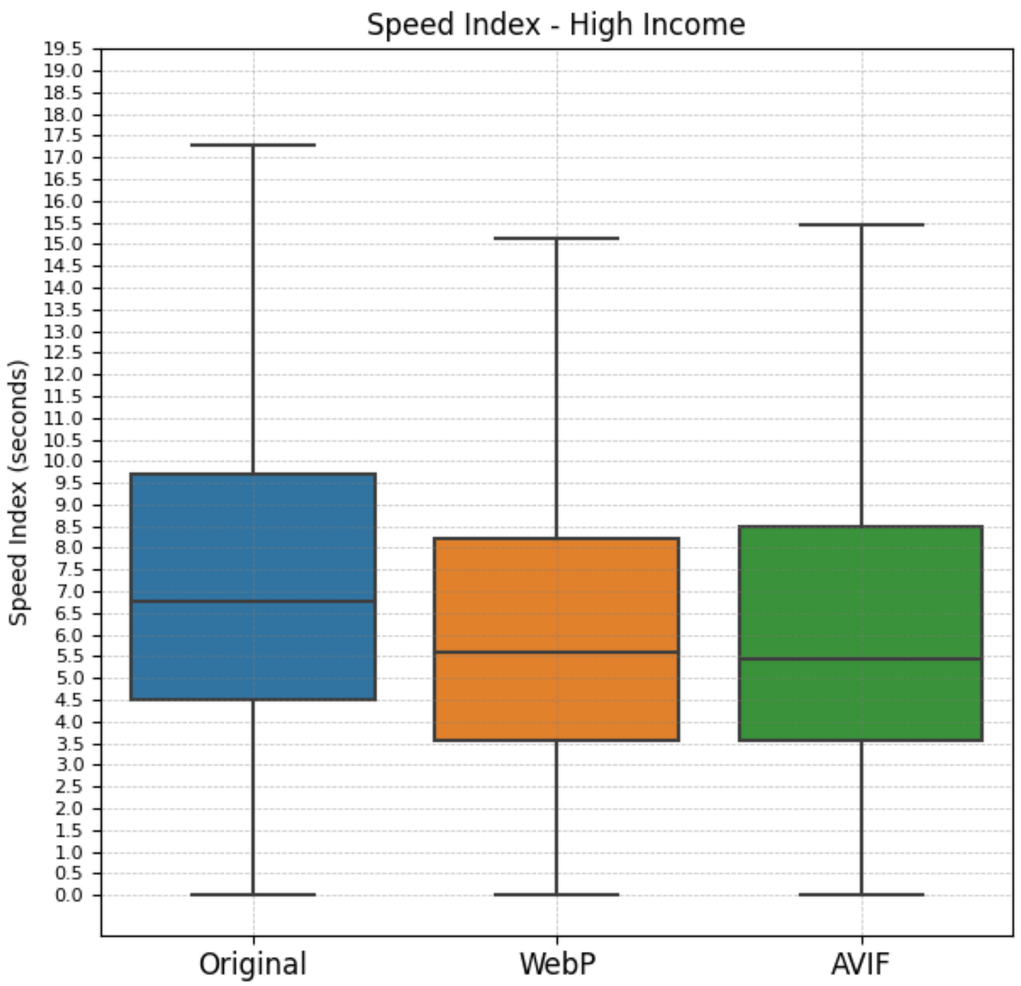
Country-wise Analysis

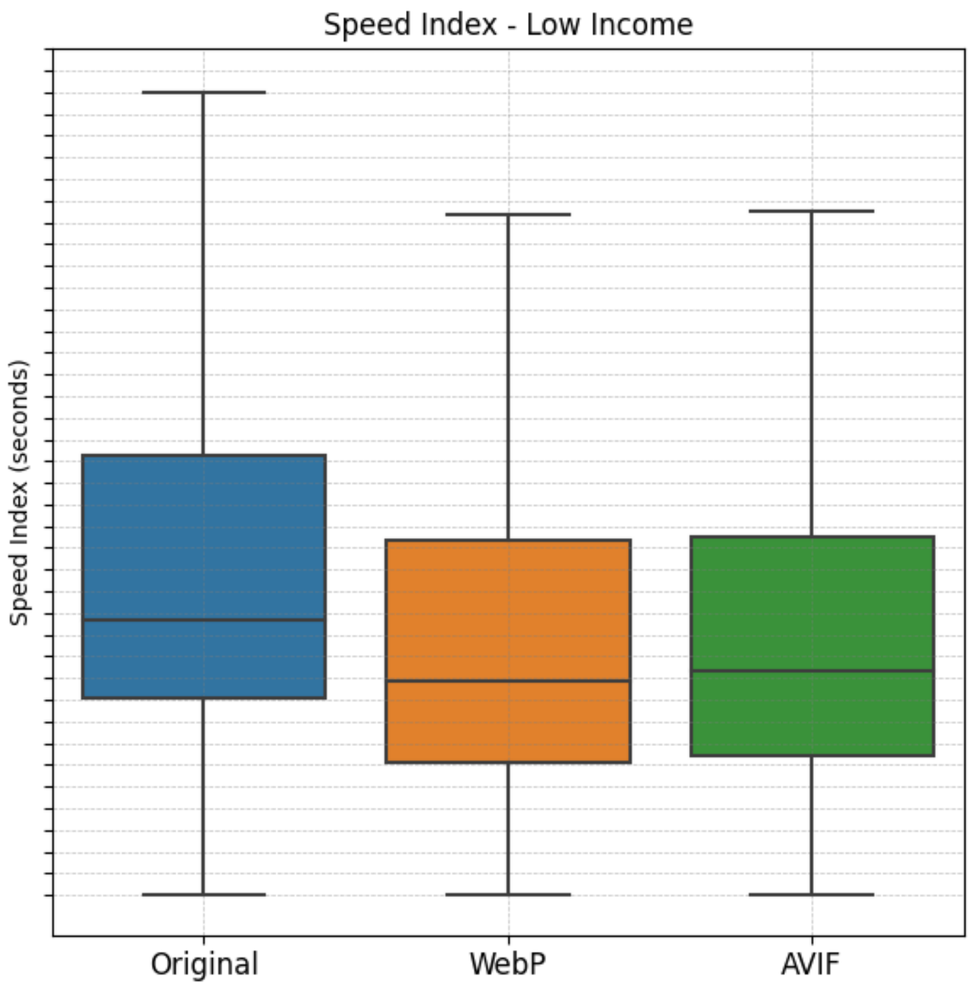


Country-wise Analysis

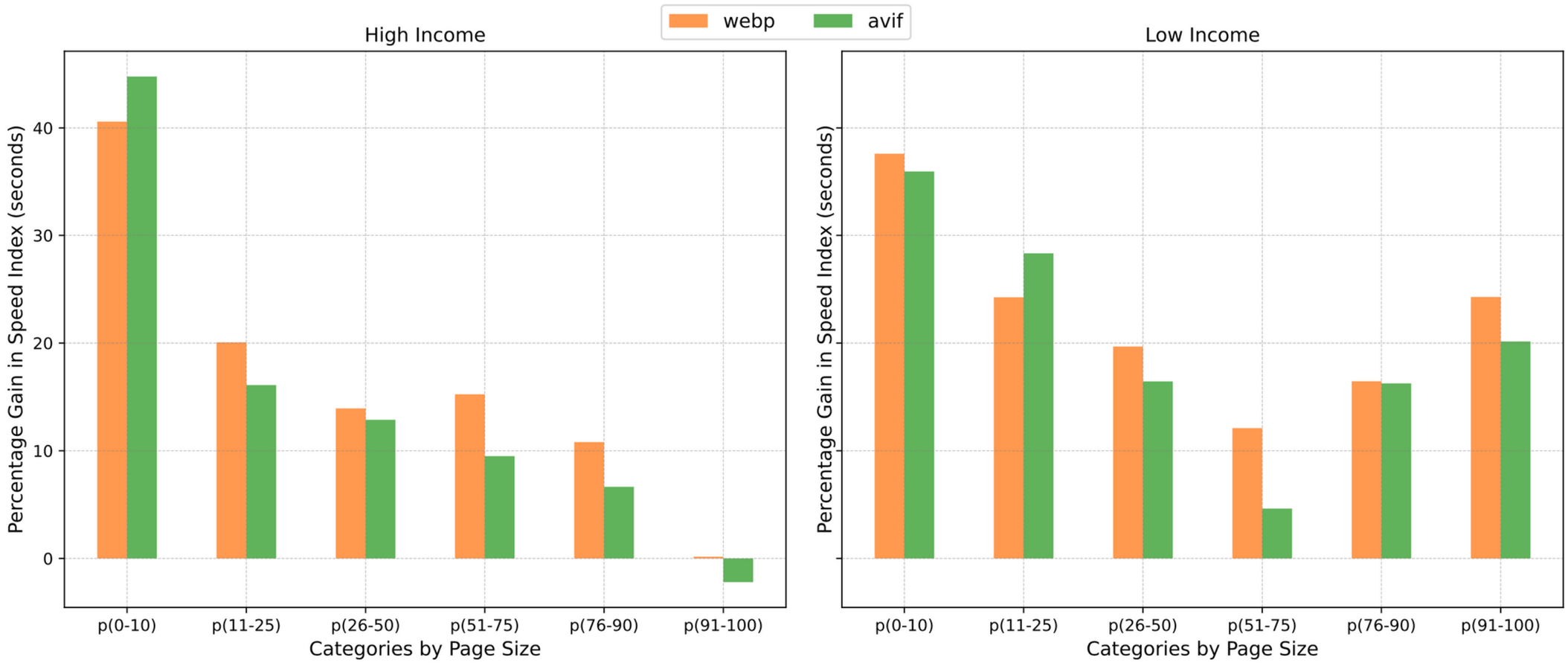


Country-wise Analysis: Data Description

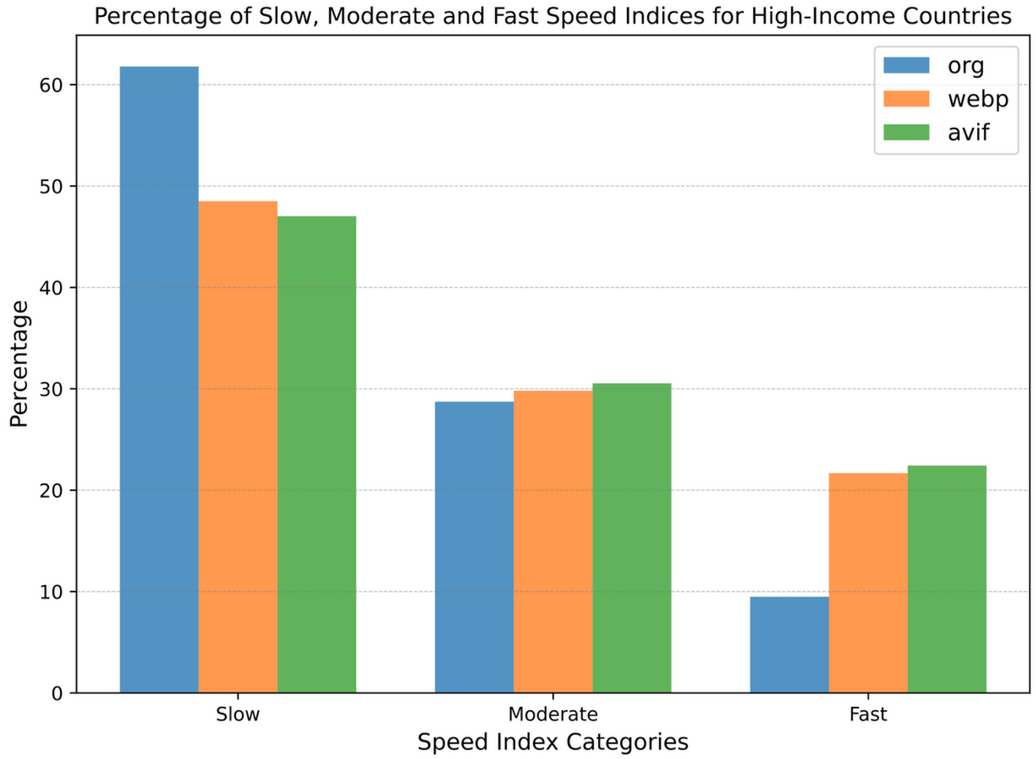


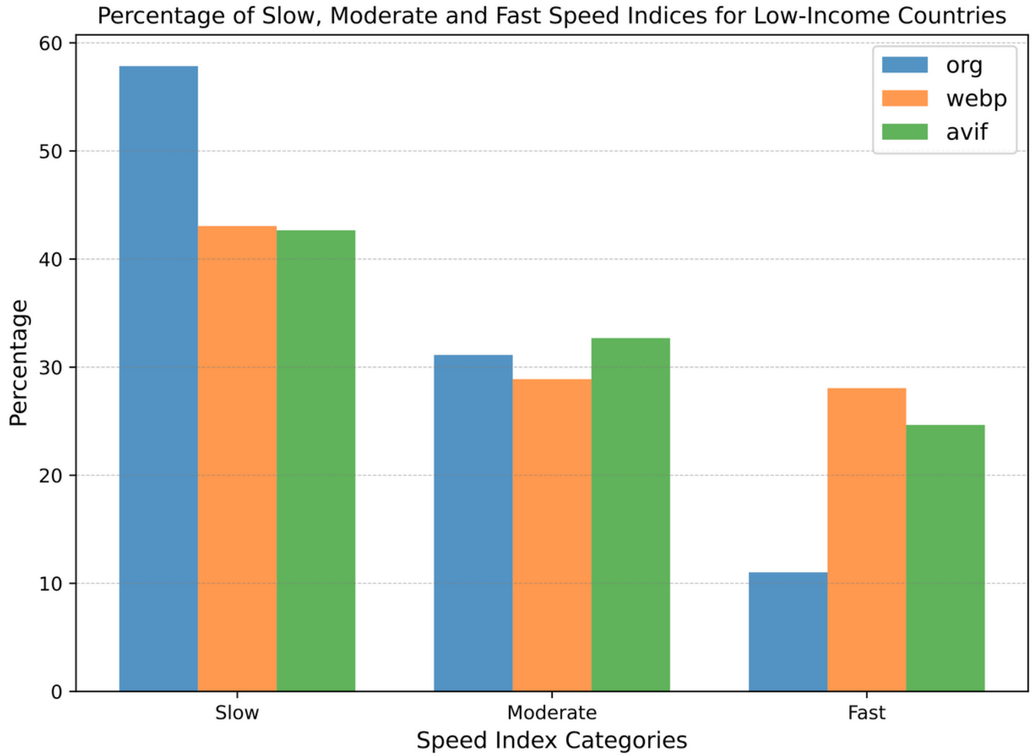


Country-wise Analysis: Data Analysis

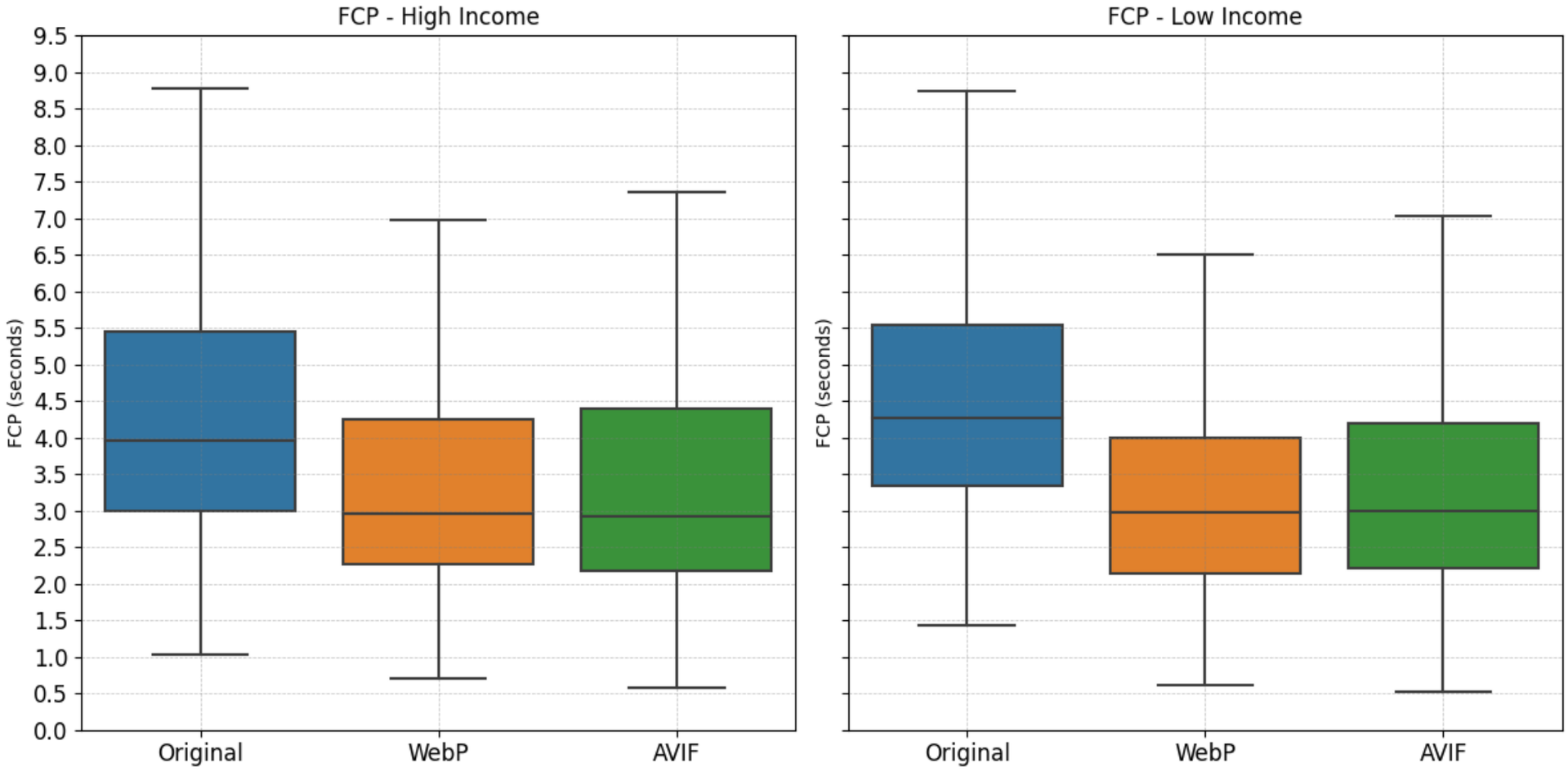


Country-wise Analysis: Data Description

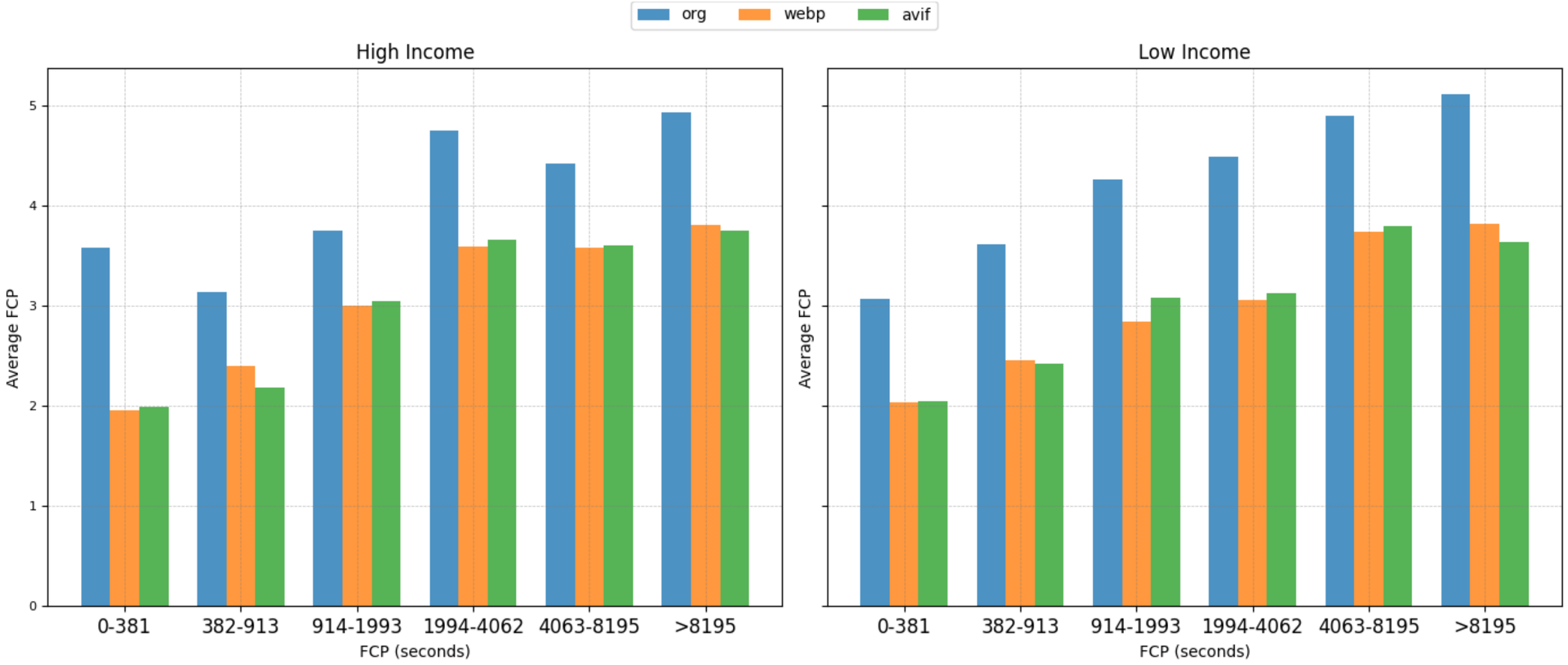




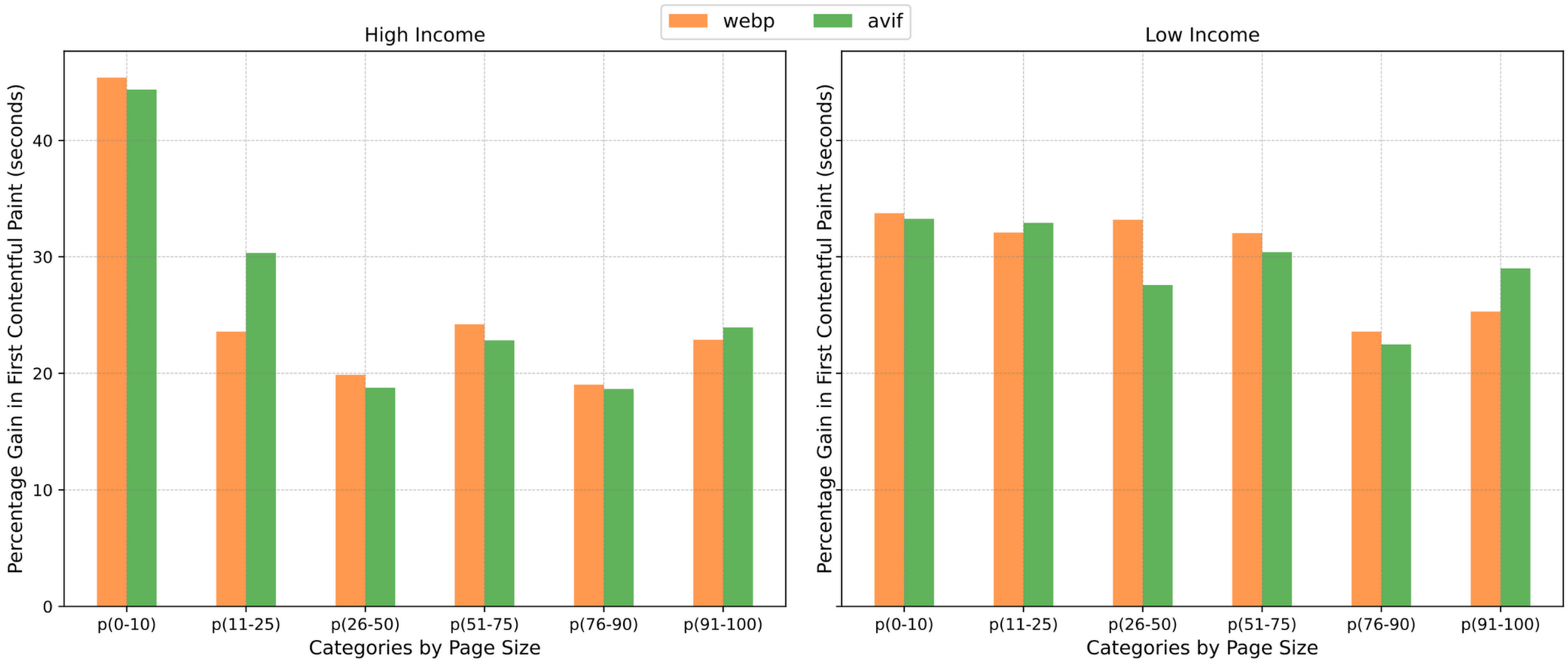
Country-wise Analysis: Data Description



Country-wise Analysis: Data Analysis



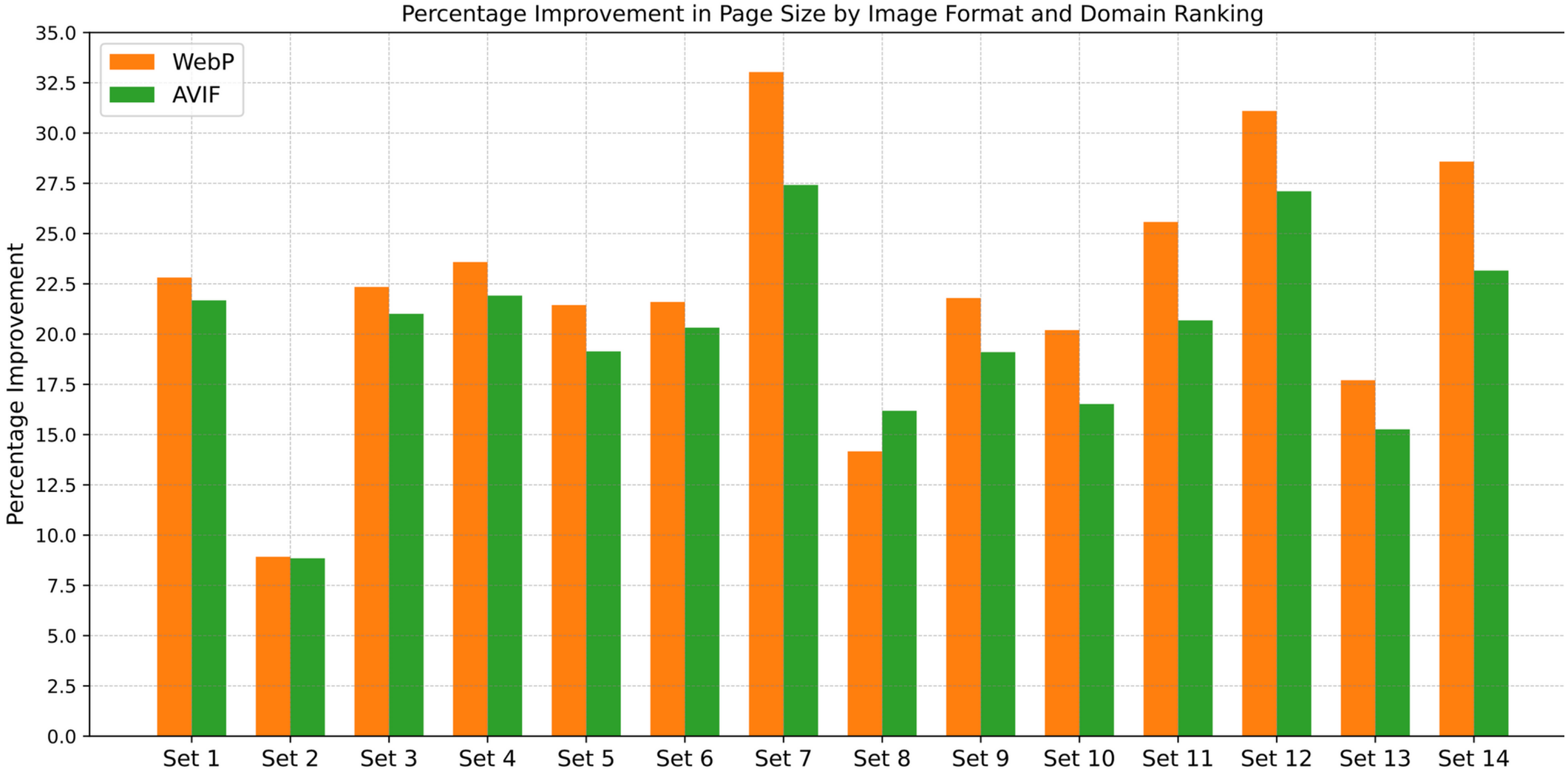
Country-wise Analysis: Data Analysis



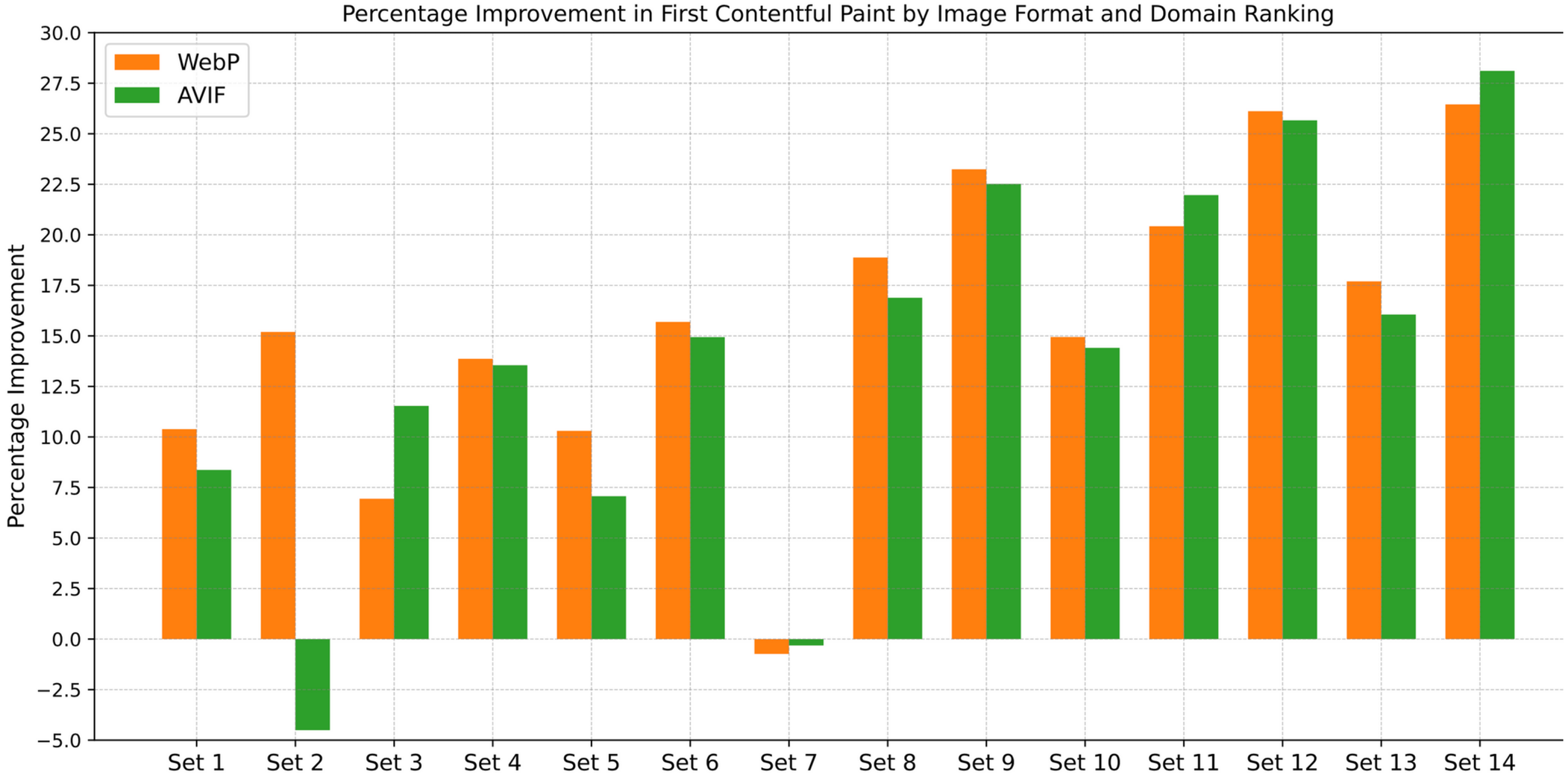
Representative Set - URL Selection

r_length	s_rank	e_rank	n_picked	% picked
100	1	100	100	100
200	101	300	100	50
400	301	700	100	25
800	701	1500	100	12.5
1600	1501	3100	100	6.25
3200	3101	6300	100	3.13
6400	6301	12700	100	1.56
12800	12701	25500	100	0.78
25600	25501	51100	100	0.39
51200	51101	102300	100	0.2
102400	102301	204700	100	0.1
204800	204701	366410	100	0.05
		Total	1200	

Representative Set - Results



Representative Set - Results



Representative Set - Results

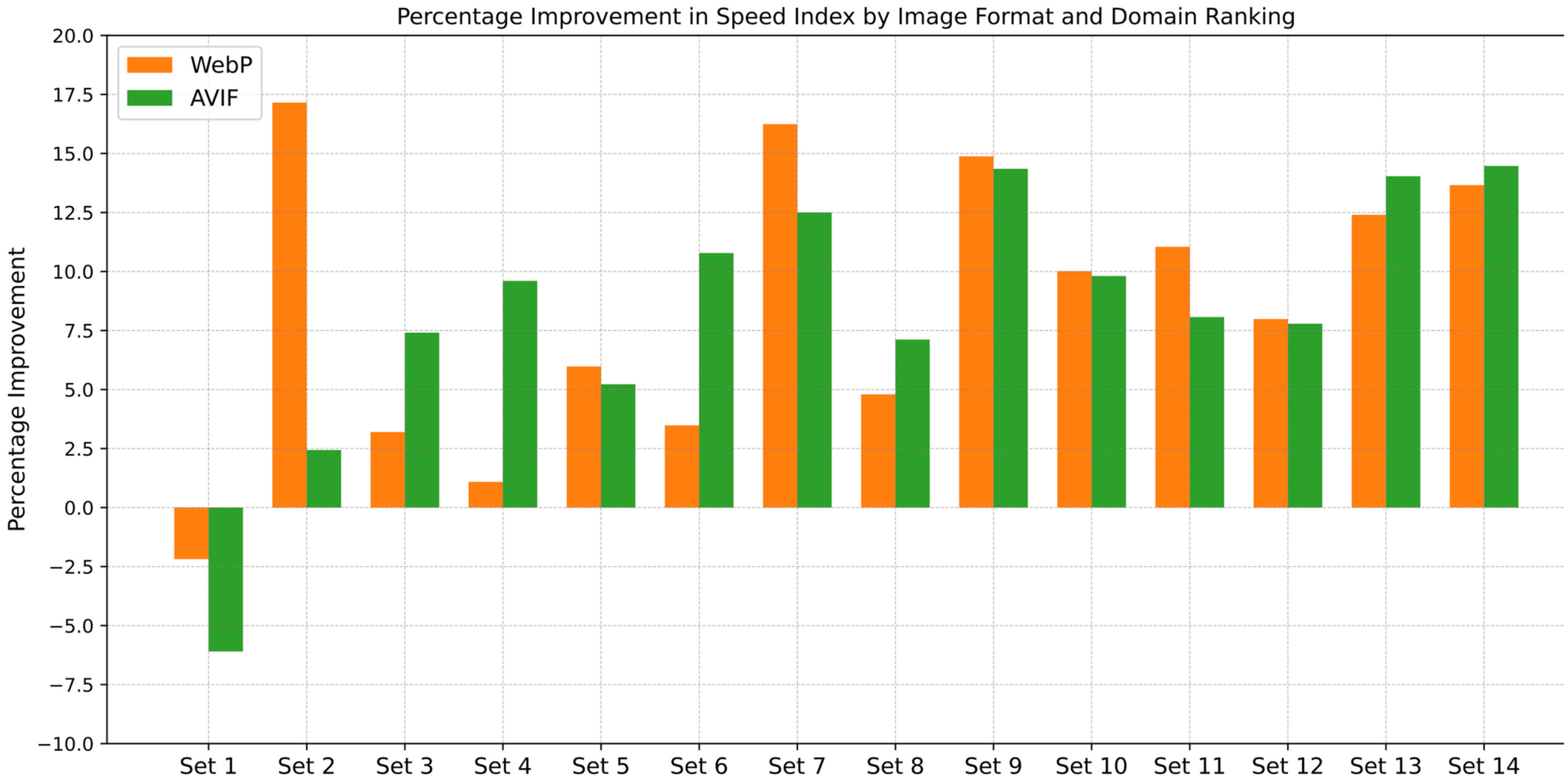


Image Level Analysis - Results

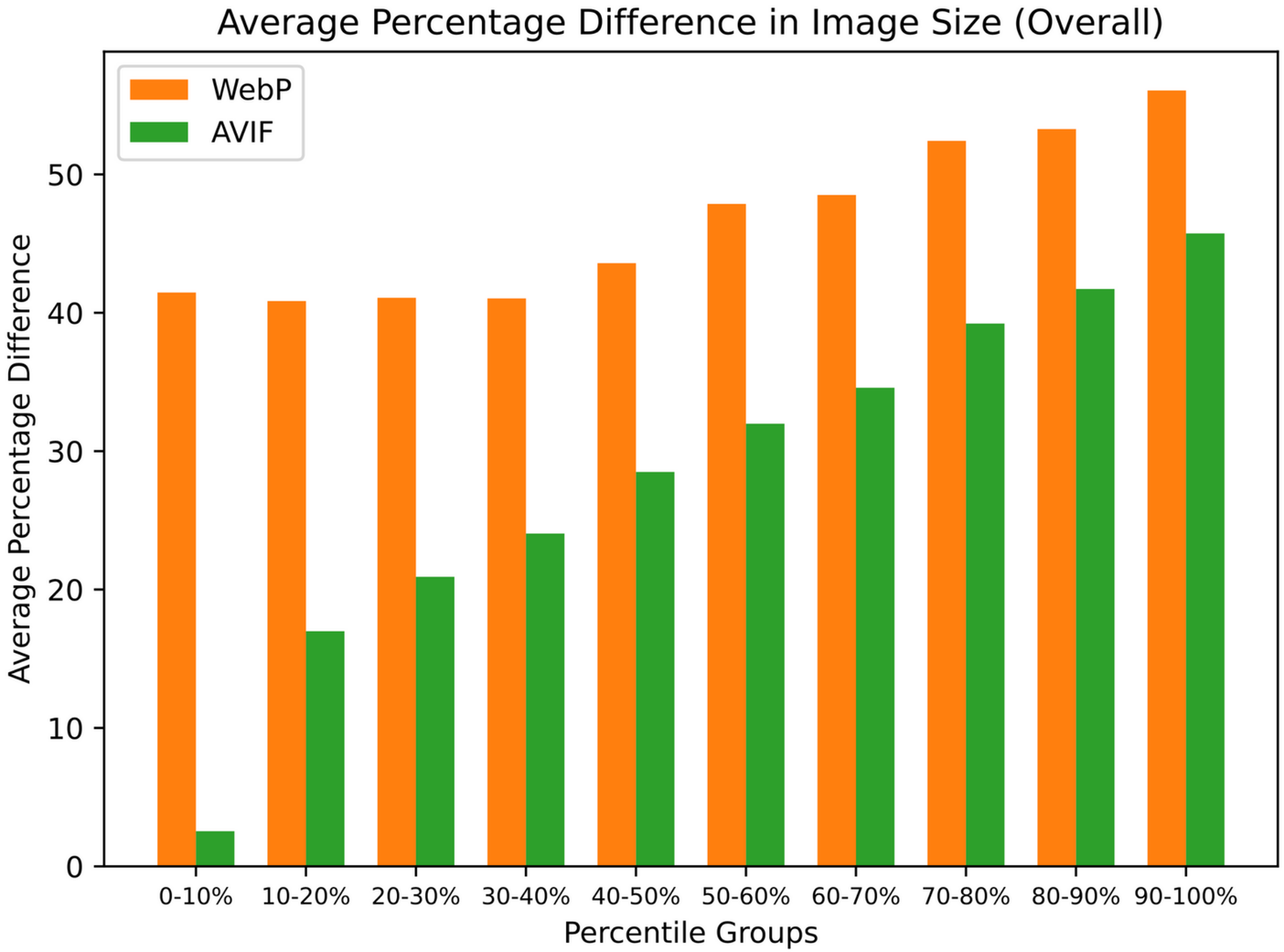


Image Level Analysis - Results

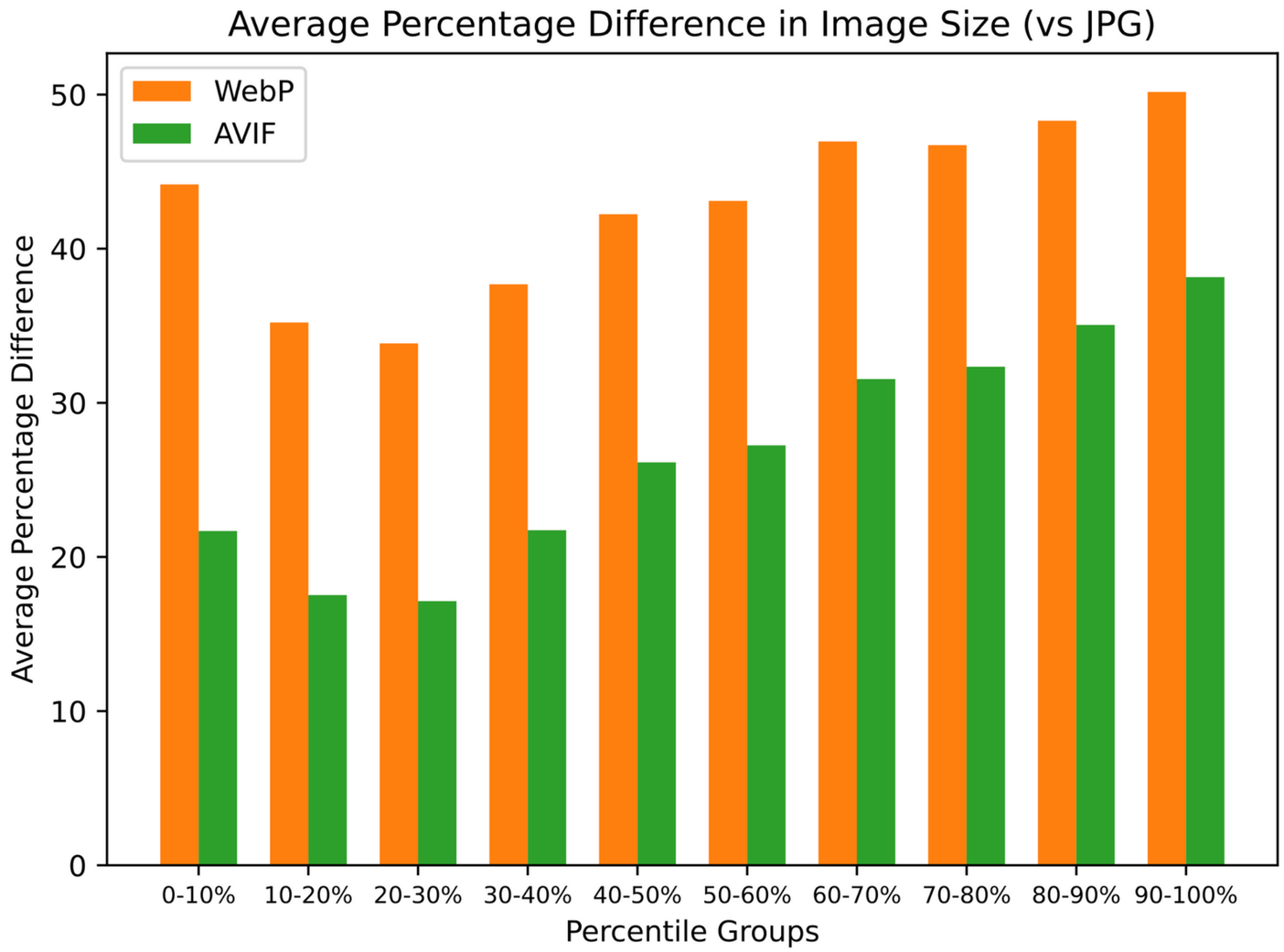
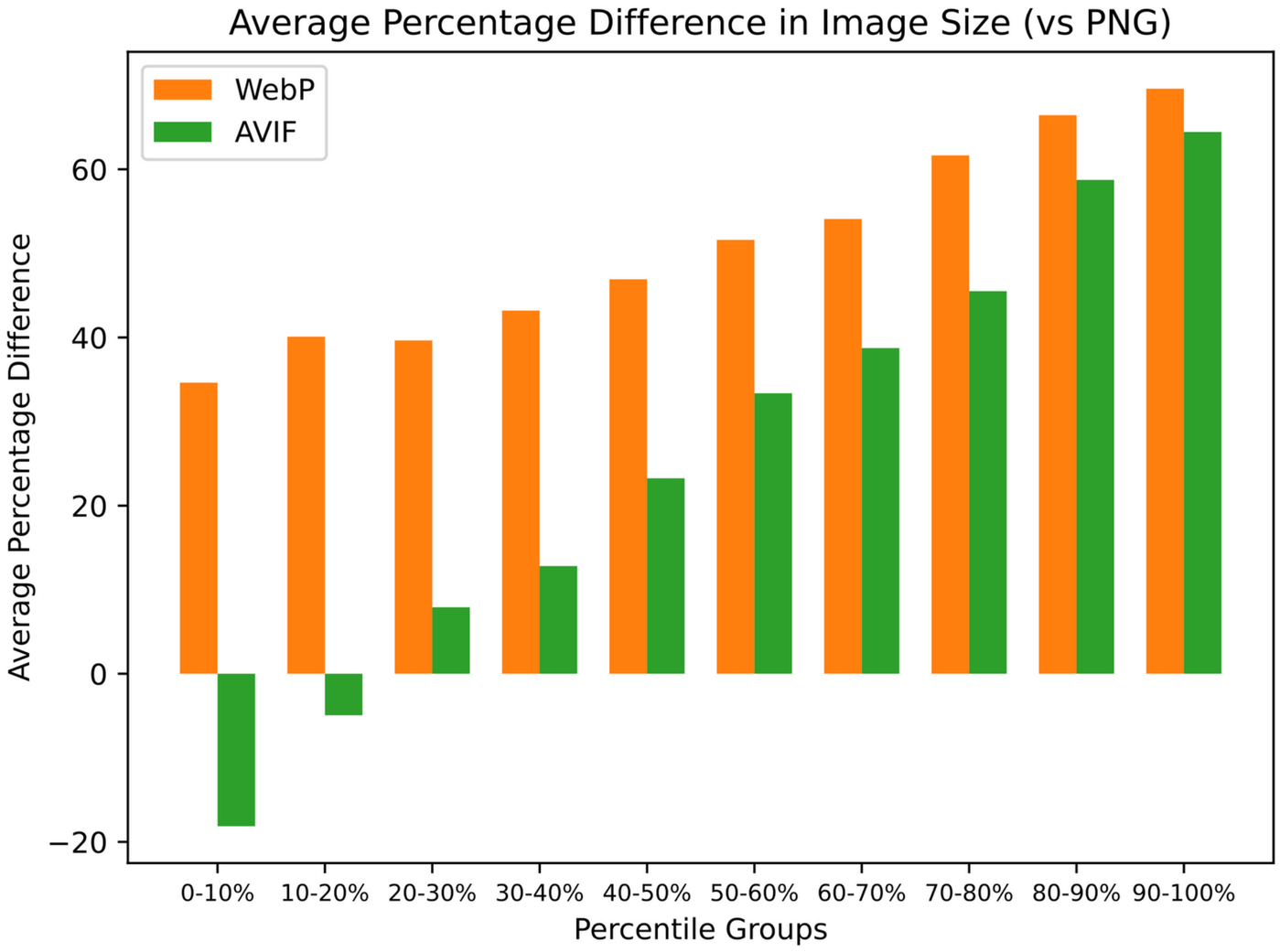


Image Level Analysis - Results



png deciles

1.0409947359625669
1.6620780414438503
2.5038913292589764
3.6573354839094727
5.255028289724981
7.73312762605042
11.723993894432773
18.7736255013369
32.96455727177234
65.31742291468632

Total Org Size: 3.04GB
Total AVIF Size: 1.33GB (down 56%)
Total WEBP Size: 1.06GB (down 65%)

Areas of Further Research

The following directions may be pursued for further research:

- Category-wise analysis (news vs sports, etc)
- An Analysis of the processing power and time required to decode WebP, AVIF and JPG/PNG images.
- Repeating the analysis on a high-end mobile device.
- Repeating the analysis with varying network conditions and types.