Identifying Dust-Obscured Galaxy Candidates in SDSS Using a Colour–Extinction Approach

[1]Yakubu, M., *[2]Vwavware, J. O., [3]Ojobeagu, A. O., & [2]Ohwofosirai, A.

[1]Department of Industrial Physics, Enugu State University of Science and Technology, Enugu, Nigeria.
[2]Department of Physics, Dennis Osadebay University, Asaba, Delta State, Nigeria.
[3]Department of Industrial and Medical Physics, David Umahi Federal University of Health Sciences, Uburu, Ebonyi State, Nigeria.

*Corresponding author email: oruaode.vwavware@dou.edu.ng.

Abstract
We present a systematic analysis of dust-obscured galaxy (DOG) candidates from a parent sample of ~50,000 galaxies in the Sloan Digital Sky Survey (SDSS). Using a selection method based on color and extinction thresholds, we identified 12,209 galaxies that represent a statistically significant population of dusty systems. These DOG candidates exhibit properties that are systematically different from the general SDSS galaxy population, including a higher mean redshift ($z \approx 0.35$). Their average r-band extinction ($Ar \approx 0.145$) significantly exceeds that of the broader sample ($Ar \approx 0.09$), confirming enhanced dust attenuation. Our analysis of correlations between redshift and color indices reveals important evolutionary signatures. While color and redshift show a strong positive correlation, a notable negative correlation between extinction and redshift ($\rho \approx -0.59$) suggests a geometric or physical evolution of dust. This study provides a foundational, purely optical baseline for isolating dusty galaxies in SDSS and sets the stage for future multi-wavelength research.

Keywords: Correlation, Dust, Extinction, Photometric Properties, Redshift

Introduction
Dust-obscured galaxies (DOGs) represent a crucial population for understanding galaxy evolution, as dust extinction significantly biases traditional optical selection methods. Dust grains absorb and scatter stellar radiation, especially at ultraviolet and optical wavelengths, causing reddening and the underrepresentation of dusty systems in optical surveys. This extinction effect alters observed galaxy properties and limits our ability to trace the total star formation activity in the universe, as a substantial fraction of galaxies remain hidden behind their own dust (Calzetti, 2001; Riguccini et al., 2011). Observational evidence shows that color selection techniques frequently miss significant numbers of star-forming galaxies at redshifts around z ~ 2–3. For instance, the BzK and BM/BX selection criteria can overlook up to 50% of mid-infrared sources due to dust reddening (Riguccini et al., 2011). DOGs exhibit a composite nature, where fainter 24 μm sources are typically dominated by star formation, while brighter sources display increasing active galactic nucleus (AGN) contributions to their total infrared luminosity (Riguccini et al., 2015). Moreover, dust opacity removes half or more of the stellar energy from the UV–optical budget, profoundly influencing our understanding of galaxy evolution and hindering accurate obscuration estimates at high redshift (Calzetti, 2001).

Theoretical models further support the evolutionary significance of DOGs. Simulations suggest that luminous DOGs originate from gas-rich mergers occurring in massive dark matter halos, characterized by extreme star formation rates of approximately 500–1000 M⊙/yr and strong AGN activity. These systems are believed to evolve from starburst-dominated to AGN-dominated phases during galactic coalescence, marking an important transitional stage in galaxy evolution (Narayanan et al., 2010). Consequently, studying these dusty populations provides critical insights into the physical mechanisms driving starburst–AGN interplay and feedback in galaxies.

Empirical investigations have quantified the importance and prevalence of dusty star-forming galaxies across different wavelengths. Davoodi et al. (2006) found that about 18% of their SDSS–SWIRE sample exhibited red optical colors with infrared excess, contributing nearly 28% of the total infrared luminosity density despite representing only 13% of the galaxy population. Rosario et al. (2016) showed that far-infrared selection techniques reveal large populations of cold, dust-dominated galaxies that optical surveys fail to detect, emphasizing the indispensability of infrared data for a complete

census of star formation. Similarly, Riguccini et al. (2011) demonstrated that traditional optical selection criteria, such as BzK and BM/BX, can miss up to 50% of mid-infrared-selected sources due to severe dust reddening effects. Casey et al. (2014) further observed that the brightest dusty star-forming galaxies can be completely optically obscured despite exhibiting extreme star formation rates, underscoring the fundamental limitation of optical surveys in capturing the full extent of cosmic star formation.

Despite these advances, a methodological gap persists in identifying dusty galaxies using purely optical data. Most previous studies have relied on infrared or multi-wavelength observations to infer dust attenuation, leaving a lack of standardized optical-only selection methods. The aim of this study is to develop and validate a reproducible, purely optical methodology for identifying dust-obscured galaxy (DOG) candidates in the Sloan Digital Sky Survey (SDSS). To achieve this aim, the following objectives were pursued:

a.   To establish robust selection criteria based on photometric color indices (g–r, r–i) and r-band extinction (A_r) thresholds.

b.   To apply these criteria to a large, clean SDSS sample (~50,000 galaxies) to construct a well-defined catalogue of DOG candidates.

c.   To statistically characterize the photometric, redshift, and extinction properties of the selected candidates and compare them with the general galaxy population.

d.   To analyses correlation patterns among redshift, color, and extinction to infer signatures of dust evolution.

While previous works such as Davoodi et al. (2006) combined optical and infrared data, our approach is distinct in its exclusive reliance on a purely optical baseline. The novelty of this work lies in offering one of the first systematic, purely optical baselines for dusty galaxy identification within SDSS, thereby providing a foundation for future multi-wavelength analyses to better constrain the role of dust in cosmic star formation.

Methods and Materials
Data acquisition and cleaning
The dataset for this study was obtained from the Sloan Digital Sky Survey (SDSS) Data Release 17 (DR17), which provides both photometric and spectroscopic data for millions of galaxies. The use of this large, publicly available astronomical survey aligns with modern approaches in data-driven astronomy, as exemplified by the methodologies discussed in Vwavware et al. (2025). Our parent sample consists of approximately 50,000 galaxies with measured magnitudes in the five SDSS bands (u, g, r, i, z), spectroscopic redshifts (z), and r-band extinction values (Ar). Prior to analysis, the dataset was carefully filtered to ensure a clean and reliable sample. Duplicate entries and objects classified as stars or quasars were excluded using the SDSS photometric and classification flags, specifically selecting only those with class = GALAXY.

To further improve data quality, we applied the clean flag and removed galaxies with high redshift errors (z_err > 0.1) or poor photometric quality (psf_flux_r < 0). This process produced a robust and reliable dataset suitable for analysis.

To isolate dusty galaxy candidates, a dual selection approach combining photometric color indices and extinction thresholds was applied. Since dust reddening systematically affects galaxy colors, we selected galaxies satisfying the criteria (g–r > 1.71) and (r–i > 0.65) values corresponding to the 75th percentile of the parent sample. This cutoff isolates the reddest quartile of galaxies while maintaining statistical robustness. Additionally, to ensure the presence of significant dust attenuation, we imposed a requirement of Ar > 0.16, defined as the mean plus one standard deviation of the extinction distribution. These combined criteria yielded 12,209 dusty galaxy candidates, representing roughly one-quarter of the parent sample.

Data analysis and visualization
After selection, a comprehensive statistical and visual analysis was performed on both the dusty galaxy candidates and the full parent sample. Descriptive statistics including means, standard deviations, and quartile ranges were computed for parameters such as redshift, magnitudes, color indices, and extinction values, enabling direct comparisons between the two populations.

To explore the distributions of key variables, histograms of redshift, color indices, and extinction were generated. These visualizations confirmed that dusty galaxies are generally redder and tend to occupy higher redshifts compared to the general population.

A correlation matrix was constructed to quantify relationships among major parameters, particularly focusing on redshift, color indices (g–r, r–i), and extinction (Ar). The Pearson correlation coefficients revealed strong positive correlations between redshift and color, and a significant negative correlation between redshift and extinction. Finally, scatter plots of redshift versus color and redshift versus Ar were generated to visually confirm these statistical relationships. The resulting trends clearly demonstrated that dustier galaxies exhibit redder colors and are typically located at moderately higher redshifts, consistent with the expectations of dust-reddened galaxy populations.

Results

The application of our color extinction selection criteria to the parent sample of ~50,000 SDSS galaxies yielded 12,209 dusty galaxy candidates, representing 24.3% of the total. A statistical comparison confirms that this subsample is systematically distinct from the general galaxy population in both redshift and dust content.

Comparative properties of dusty candidates: The dusty candidates are, on average, at higher redshifts, are redder in color, and exhibit greater dust attenuation than the parent sample. The key comparative statistics are summarized in Table 1.

Table 1: Comparison of Descriptive Statistics for Key Parameters

Parameter	Sample	Count (N)	Mean	Standard Deviation (σ)
Redshift (z)	Parent Sample	49,975	0.254	0.121
	Dusty Candidates	12,209	0.346	0.116
Color Index (g–r)	Parent Sample	49,975	1.343	0.434
	Dusty Candidates	12,209	1.639	0.388
r-band Extinction (A_r)	Parent Sample	49,975	0.090	0.072
	Dusty Candidates	12,209	0.144	0.115

Figure 1a: Redshift distribution of dusty galaxy candidates compared to the parent SDSS sample. The dusty
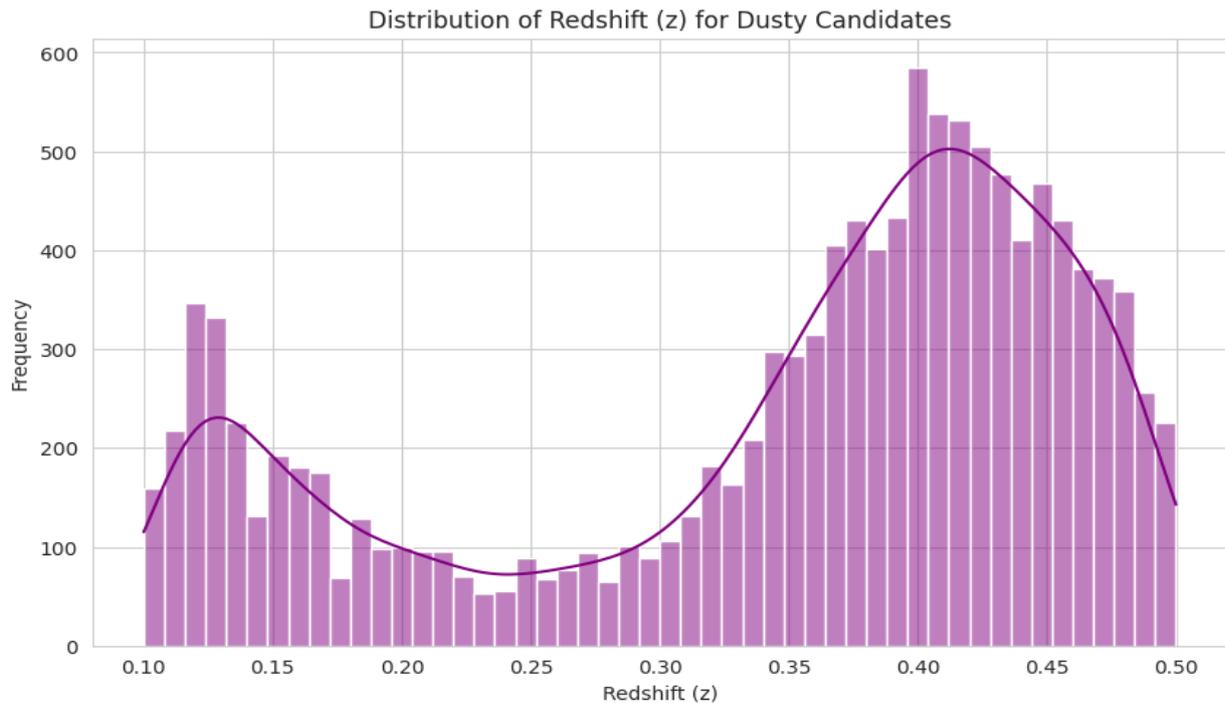
candidates show a distribution shifted toward higher redshifts with peaks at $z \approx 0.15$ and $z \approx 0.4$. This distribution is clearly shifted towards higher redshifts compared to the parent sample (Figure 1), confirming that our selection method preferentially identifies more distant systems.
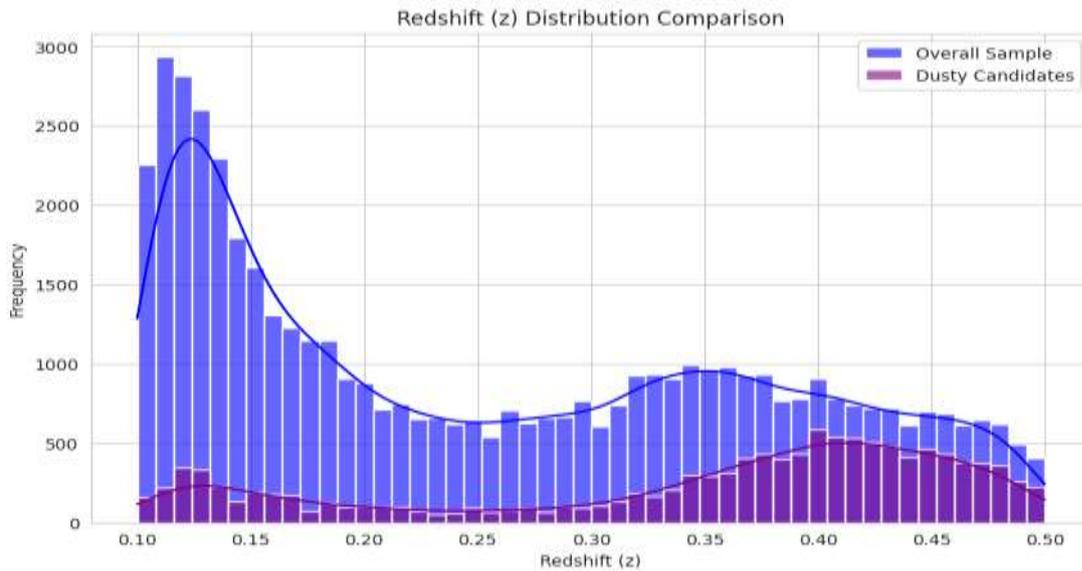
Figure 1b: Bimodal redshift distribution of dusty galaxy candidates. The parent sample (gray, inset) peaks at $z \approx 0.25$, while dusty candidates show a broader distribution with a dominant
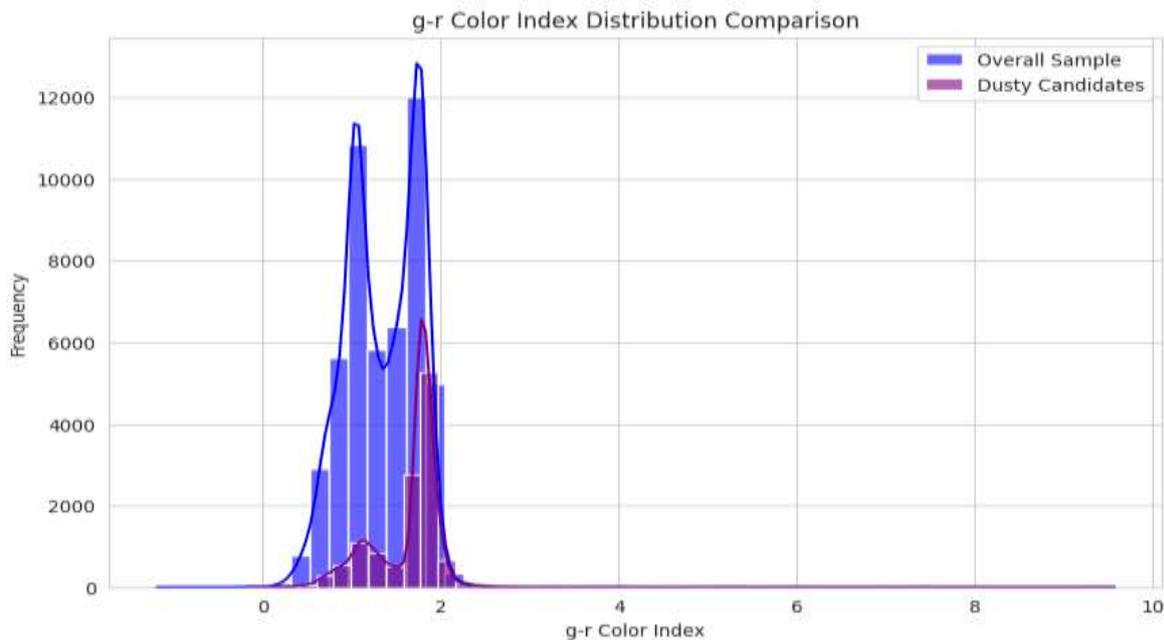peak at $z \approx 0.40$.

Figure 2a: Distribution of g–r color index for the parent sample (gray) and dusty candidates (red). Dusty candidates populate exclusively the redder peak (g–r $\approx$ 1.8–2.0).
The dusty candidates are concentrated almost exclusively in the redder peak (g–r $\approx$ 1.8–2.0), indicating their intrinsically redder colors due to dust extinction.
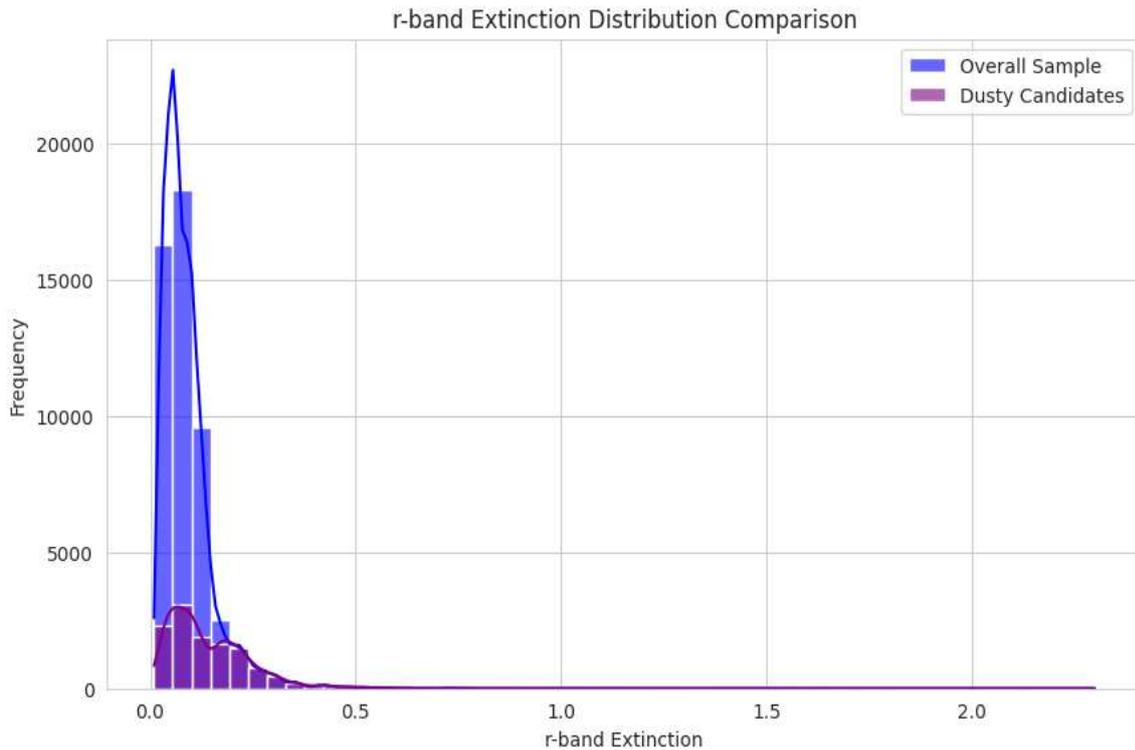
Figure 2b: Distribution of r-band extinction (A_r) for the parent sample (gray) and dusty candidates (red). Dusty candidates show significantly higher extinction values.

The overall sample shows a sharp peak near $A\_r \approx 0.0$–0.1, indicating minimal extinction. In contrast, the dusty candidates exhibit a broader distribution, with a significant fraction extending to $A\_r \approx 0.3$–0.5, confirming that they experience higher extinction on average.

Correlation analysis: The correlation matrix for the dusty candidates (Table 2) reveals the fundamental relationships between their properties. We find strong positive correlations between redshift and the color indices u-g ($\rho \approx 0.76$), g-r ($\rho \approx 0.75$), and r-i ($\rho \approx 0.76$). This indicates that more distant candidates are systematically redder.

Most notably, we observe a significant negative correlation between redshift and r-band extinction ($\rho \approx -0.59$). This trend suggests an evolution in the nature or distribution of dust, where higher-redshift dusty galaxies in our sample are associated with lower measured extinction values, potentially pointing to a geometric or physical evolution of dust with cosmic time.

Table 2: Correlation Matrix - Dusty Candidates (12,209 Galaxies)

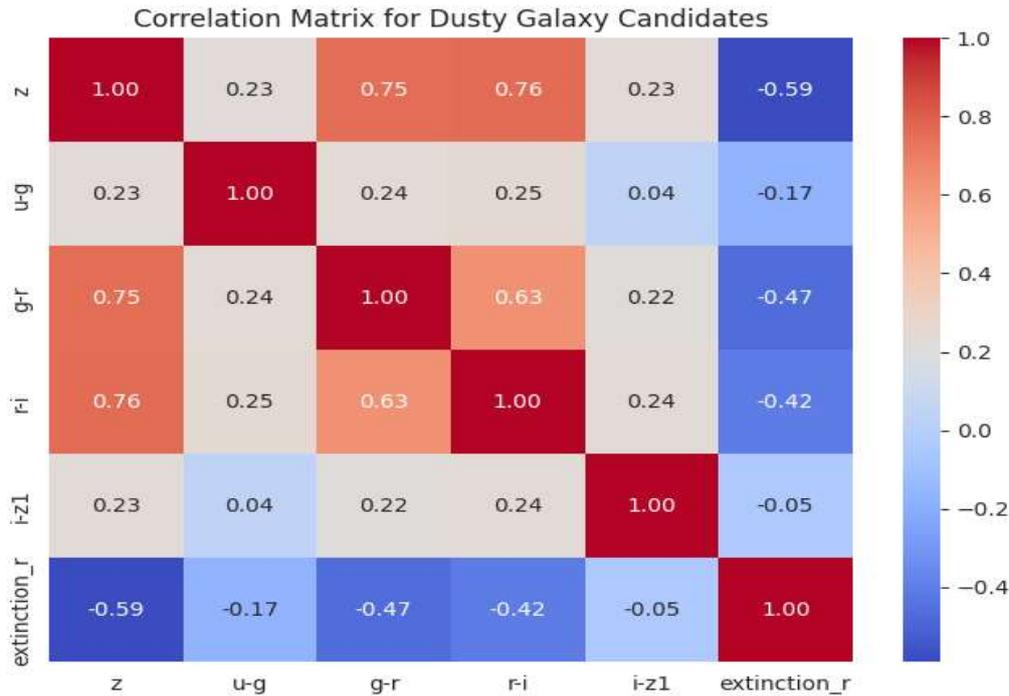| Parameter	Z	u–g	g–r	r–i	i–z_1	A_r (extinction)
Z	1.000	0.762	0.752	0.763	0.233	–0.591
u–g	0.762	1.000	0.239	0.249	0.042	–0.173
g–r	0.752	0.239	1.000	0.629	0.222	–0.472
r–i	0.763	0.249	0.629	1.000	0.237	–0.416
i–z_1	0.233	0.042	0.222	0.237	1.000	–0.047
A_r (extinction)	–0.591	–0.173	–0.472	–0.416	–0.047	1.000

Figure 3: Correlation matrix for the 12,209 dusty galaxy candidates, showing Pearson correlation coefficients between redshift (z), color indices, and extinction (A_r).

Figure 3, Redshift shows strong positive correlations with color indices: u–g ($\rho = 0.76$), g–r ($\rho = 0.75$), r–i ($\rho = 0.76$), confirming that more distant galaxies are redder. A notable negative correlation exists between redshift and r-band extinction ($\rho = -0.59$), indicating lower measured extinction at higher z. Other parameters such as i–z_1 ($\rho = 0.23$) and extinction with i–z_1 ($\rho = -0.05$) show only weak relationships.
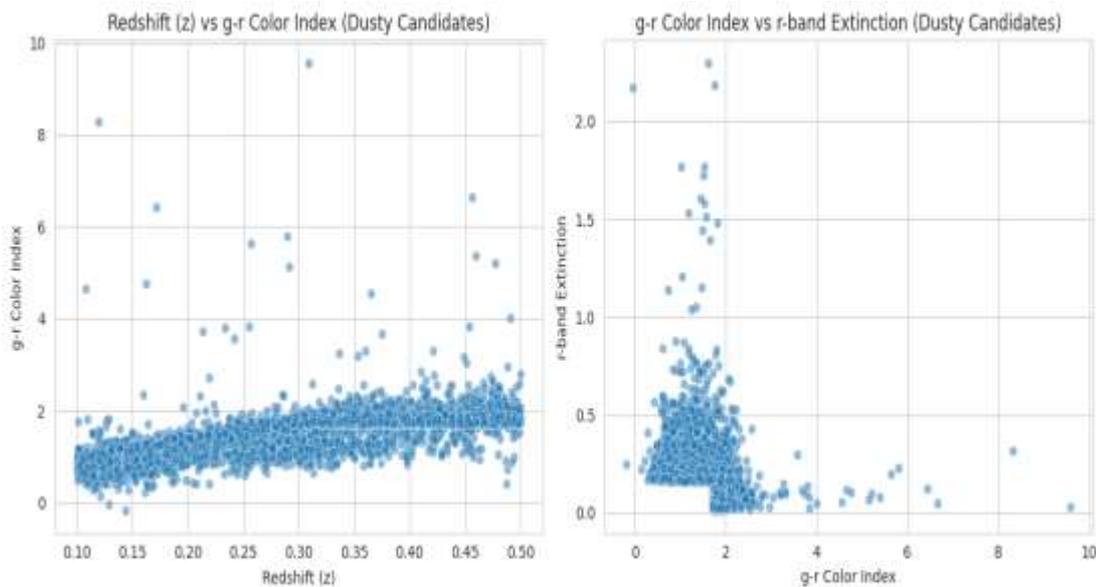

Figure 4: Scatter plots for dusty candidates. Left: Redshift versus g–r color ($\rho \approx 0.75$). Right: g–r color versus r-band extinction ($\rho \approx -0.47$)

Figure 4 consists of two scatter plots for dusty candidates, whose light is obscured by dust. The left plot shows that as an object's redshift (distance) increases, it becomes redder, a result of the universe's expansion. The right plot demonstrates that a significant portion of this red color is due to dust extinction, which preferentially absorbs and scatters bluer light.

In the left panel, redshift correlates positively with g–r color ($\rho \approx 0.75$), showing that higher-z dusty candidates are systematically redder. In the right panel, g–r color correlates negatively with r-band extinction ($\rho \approx -0.47$), with most galaxies clustered at g–r $\approx$ 1.5–2.0 and $A_r < 0.5$. A smaller fraction extends to g–r $>$ 3 and $A_r \approx$ 1–2, representing the reddest, most extinguished systems.
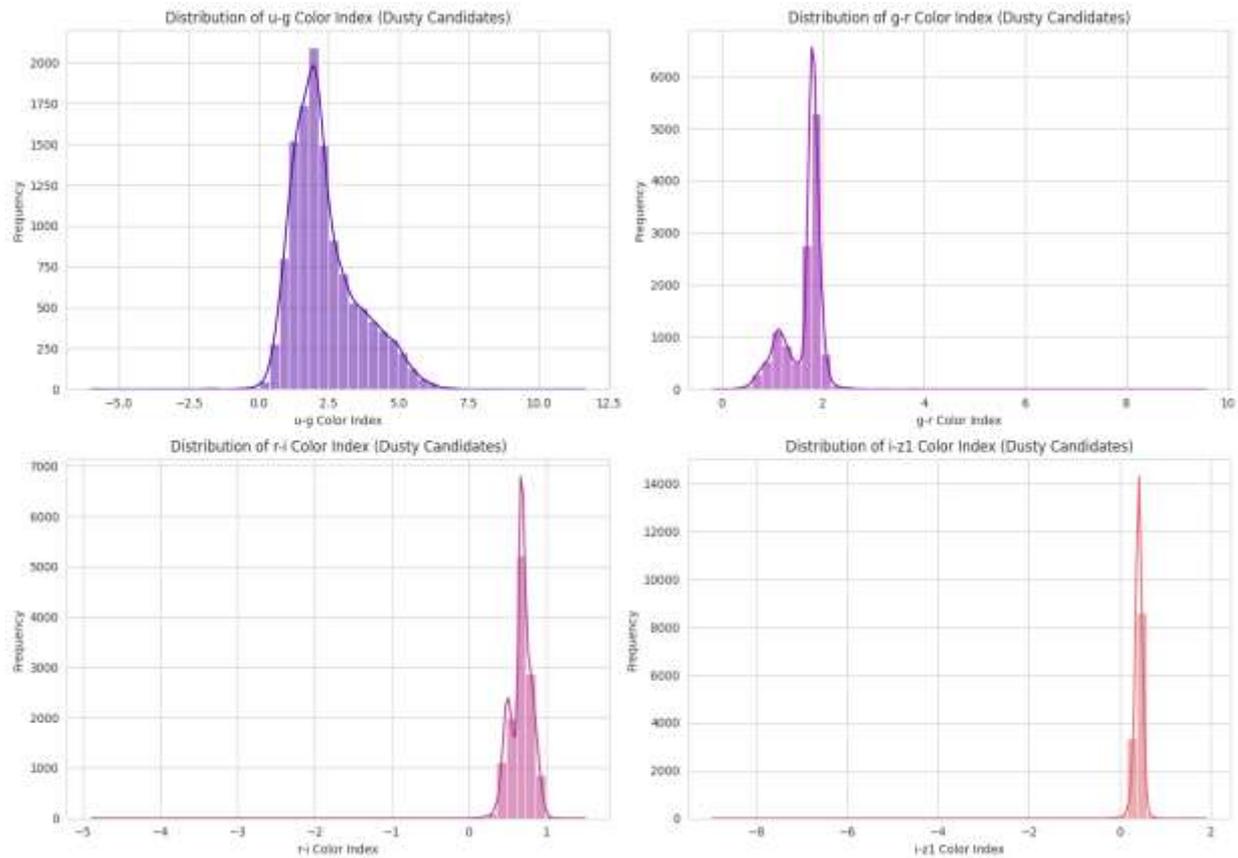

Figure 5: Histograms of color indices for dusty candidates: u–g (top left), g–r (top right), r–i (bottom left), and i–z_1 (bottom right). Note the sharp peak in i–z_1, characteristic of dust-dominated systems

u-g Color Index (Top Left): This distribution is broad and has a prominent peak around a value slightly greater than 0, with a long tail extending to the right. This suggests a wide range of properties, possibly indicating different types of dusty objects. The "u" (ultraviolet) and "g" (green) filters are sensitive to hot, young stars or active galactic nuclei.

g-r Color Index (Top Right): The distribution is more concentrated, with a strong main peak around 1.5, and a smaller secondary peak around 0.5. The "g" (green) and "r" (red) filters are commonly used to study the color of stars, and this double-peaked structure could indicate two distinct populations of objects within the sample.

r-i Color Index (Bottom Left): This histogram is highly peaked and narrow, centered near 0.5. The "r" (red) and "i" (infrared) filters are particularly useful for observing cool stars and distant galaxies. The narrow distribution suggests a very uniform set of objects in the sample, at least in this color range.

i-z1 Color Index (Bottom Right): This distribution is extremely sharp and concentrated, with a very high and narrow peak. The "i" (infrared) and "z1" (far-infrared) filters are essential for detecting very cool objects or those heavily obscured by dust. The highly peaked nature of this distribution suggests that the dusty candidates have a very consistent and similar infrared signature, which is expected for objects dominated by dust emission.

Discussion

Our analysis successfully demonstrates that a combined color–extinction approach can effectively isolate a distinct population of dust-obscured galaxy candidates from purely optical SDSS data. The identification of 12,209 candidates, representing a significant fraction (24.3%) of the parent sample, underscores that dusty systems are not a negligible minority but a substantial component of the local galaxy population, consistent with findings from earlier multi-wavelength studies (Davoodi *et al*., 2006; Riguccini *et al.,* 2011).

The systematic differences between our DOG candidates and the general galaxy population are striking. The significantly higher mean redshift (z $\approx$ 0.35) suggests that our selection method is preferentially identifying more distant, and potentially more luminous or dust-rich, systems. The bimodal structure observed in the redshift distribution (Figure 1b) is particularly

intriguing. It may indicate the presence of two distinct sub-populations of dusty galaxies: a lower-redshift group ($z \approx 0.2$) and a more dominant, higher-redshift group ($z \approx 0.4$). This could reflect different evolutionary stages, triggering mechanisms (e.g., secular evolution vs. mergers), or a selection bias where the reddest, most obscured galaxies are only detectable within a specific redshift window. The pronounced bimodality in the g-r color distribution of the parent sample, with our candidates exclusively populating the red peak (Figure 2a), provides strong, purely optical validation that our criteria successfully target reddened systems.

The core of our analysis lies in the correlation patterns, which reveal important signatures of dust-galaxy co-evolution. The strong positive correlations between redshift and the optical color indices (u-g, g-r, r-i; ($\rho \approx$ 0.75–0.76) are a classic signature of the color-redshift relation, where more distant galaxies appear redder due to a combination of cosmological redshift and the aging of their stellar populations. However, the notable negative correlation between redshift and r-band extinction ($\rho \approx$ -0.59) presents a more complex and compelling narrative. This trend indicates that the dustiest galaxies in our sample, as quantified by their high extinction values, are found at lower redshifts. Conversely, at higher redshifts, we find galaxies that are red in color (as shown by the positive correlations) but have systematically lower measured extinction.

This apparent paradox can be explained by several non-mutually exclusive physical and geometric scenarios:
1. Evolution of Dust Properties: The nature of dust grains themselves may evolve with cosmic time. At higher redshifts, dust grains might be smaller, have a different composition, or be more spatially diffuse, leading to less effective attenuation per unit mass (i.e., a lower extinction coefficient) while still causing significant reddening.
2. Geometric Evolution: The distribution of dust within galaxies may change. At higher redshifts, star formation and AGN activity are more prevalent, potentially driving dust into a more diffuse, extended halo or a clumpy, patchy distribution. This geometry would cause reddening without producing the strong, uniform extinction measured in nearby, more settled systems where dust resides in a centralized, screen-like disk.
3. Selection Effect: Our purely optical selection might be biased at higher redshifts. Extremely dusty galaxies (with very high A_r) could be completely obscured at optical wavelengths and thus drop out of the SDSS spectroscopic sample altogether. Therefore, the high-z end of our sample may be inherently limited to galaxies with moderate extinction, creating an artificial negative trend.

This observed evolution aligns with theoretical models that suggest luminous, dust-obscured galaxies at earlier cosmic epochs often represent major mergers in massive halos, characterized by complex dust geometries (Narayanan et al., 2010). Our findings provide an empirical, optical baseline that supports the picture of an evolving interstellar medium. The consistent and sharp peak in the i-z_i color index distribution of our candidates (Figure 5) further reinforces their nature as dust-dominated systems, as this color is particularly sensitive to the old stellar populations and dust emission that characterize such objects.

Conclusion
In this study, we have established and validated a reproducible, purely optical methodology for identifying dust-obscured galaxy candidates within the Sloan Digital Sky Survey. Through applying simple thresholds based on g–r color, r–i color, and r-band extinction, we isolated a robust sample of 12,209 galaxies that are statistically distinct from the general population in terms of redshift, color, and dust content. Our key findings are:
1. Dusty candidates reside at higher redshifts $z \approx 0.35$ compared to the parent sample $z \approx 0.25$, with a suggestive bimodal distribution.
2. They exhibit significantly redder colors and higher extinction values, confirming the efficacy of our selection criteria.
3. The correlation structure reveals a clear evolutionary signature: while galaxies are redder at higher redshifts, their measured extinction decreases, hinting at a physical or geometric evolution of dust.

The primary novelty of this work is its demonstration that a significant population of dusty galaxies can be identified using SDSS optical data alone, without reliance on infrared counterparts. This provides a crucial foundational baseline and a valuable complementary tool for future studies. The methodological gap we addressed the lack of a standardized optical-only selection method is now filled, offering a practical approach for pre-selecting interesting dusty targets from vast optical surveys. Looking forward, this catalogue of DOG candidates serves as an ideal starting point for multi-wavelength follow-up. Cross-matching with data from WISE, Herschel, or ALMA will allow for direct constraints on dust temperatures, total infrared luminosities, and star formation rates, thereby testing the physical interpretations of the redshift-extinction anti-correlation proposed here. Furthermore, spectroscopic follow-up could disentangle the nature of the bimodal populations and quantify the relative contributions of AGN and star formation to their energy output. This study thus sets the stage for a more nuanced understanding of the role of dust in shaping the observed properties of galaxies across cosmic time.

Conflict of Interest:

The authors declare that there is no conflict of interest regarding the publication of this work.

Data Availability Statement:

The SDSS DR17 dataset used in this study is publicly available at https://skyserver.sdss.org/dr17/en/tools/search/sql.aspx

References

Casey, C. M., Narayanan, D., & Cooray, A. (2014). Dusty star-forming galaxies at high redshift. *Physics Reports, 541* (2), 45–161.

Calzetti, D. (2001). The dust opacity of star-forming galaxies. Publications of the Astronomical Society of the Pacific, 113 (790), 1449 - 1485.

Davoodi, P., Pozzi, F., Oliver, S., Polletta, M., Afonso-Luis, A., Farrah, D., & Surace, J. (2006). Optical and infrared diagnostics of SDSS galaxies in the SWIRE survey. *Monthly Notices of the Royal Astronomical Society, 371* (3), 1113–1124.

Narayanan, D., Dey, A., Hayward, C. C., Cox, T. J., Bussmann, R. S., Brodwin, M., & Hernquist, L. (2010). A physical model for z~ 2 dust-obscured galaxies. *Monthly Notices of the Royal Astronomical Society, 407* (3), 1701–1720.

Riguccini, L., Le Floc'h, E., Gilbert, O., Aussel, H., Salvato, M., Capak, P., & Scoville, N. (2011). Dust-obscured star formation and the contribution of galaxies escaping UV/optical color selections at z~ 2. *Astronomy & Astrophysics, 534*, A81.

Riguccini, L., Le Floc'h, E., Mullaney, J. R., Menendez-Delmestre, K., Aussel, H., Berta, S., & Treister, E. (2015). The composite nature of Dust-Obscured Galaxies (DOGs) at z~ 2–3 in the COSMOS field–I. A far-infrared view. *Monthly Notices of the Royal Astronomical Society, 452* (1), 470–485.

Rosario, D. J., Mendel, J. T., Ellison, S. L., Lutz, D., & Trump, J. R. (2016). Local SDSS galaxies in the Herschel Stripe 82 survey: a critical assessment of optically derived star formation rates. *Monthly Notices of the Royal Astronomical Society, 457* (3), 2703–2721.

Vwavware, J. O., Mu'allim, Y., Adrian, O., & Okechukwu, O. A. (2025). The Impact of Robotic Telescopes on Time-Domain Astronomy. *Nigerian Journal of Applied Physics, 1*(1), 1-9.