Bachelor Thesis

IU International University of Applied Sciences

BSc. Data Science

AI-Driven In-Store Demographic Analytics for SME:

A Feasibility Study Using CCTV and Computer Vision to Deliver E-Commerce-Level Customer Insights

Student: Oskar Wolf

Matriculation Number: 92126079

Address: 30 Baltimore Drive, Port Elizabeth, South Africa, 6025

Supervisor: Prof. Thomas Zöller

Date of Submission: 04 December 2025

The thesis presented is based on internal, confidential data and information from the company XY. This work may not be inspected by third parties, apart from the supervisors, authorised members of the Exams Office and the Examination Board, without the explicit approval of the company and the author. Reproduction and publication of the thesis—even in parts—without express permission is not permitted.

Acknowledgement

I would like to extend my sincere appreciation to my supervisor, Professor Thomas Zöller, for his guidance, support, and expertise throughout this thesis.

Moreover, I acknowledge my lecturers as the best motivators during the course of my studies: Prof. Bertram Taetz, Prof. Frank Passing, Prof. Veronica Mas, Dr. Tolga Ülkü , and Lecturer Hashem Zarafat. I'd also like to thank Dean, who was my mathematics tutor in my first year, for his early support.

I appreciate the owner of the participating SME for granting access to the system, as well as the maintenance team that installed the camera. I’m also grateful to the café staff who assisted customers in completing the surveys.

I would like to give special acknowledgement to the open-source communities that have contributed to this project: the DeepFace developers Sefik Serengil and Alper Ozpinar, Ultralytics, OpenCV, Hugging Face, and the Ubiquiti developer community.

My thanks to my fiancée Nina, my parents, my brothers, my sister, and my uncles Rüdiger and Lorenz. I also want to thank Rüdiger for assisting me with my technical blog.

I also want to thank my grandmother for all her financial assistance and unwavering belief in my studies.

Abstract

This study explores an innovative method to empower small and medium-sized enterprises (SMEs) by providing data-driven insights that are typically accessible only to larger companies. SMEs often face challenges such as limited access to customer demographics and advanced analytics compared to e-commerce platforms. This thesis examines the feasibility and application of an affordable, privacy-preserving system that transforms point-of-sale (POS) data into demographic insights using existing CCTV infrastructure. By integrating advanced computer vision algorithms, the system correlates transaction data with demographic profiles, offering critical insights into purchasing behaviour. Analysis of a two-day dataset with nearly 2,000 transactions demonstrates the reliability of gender classification and addresses challenges such as lighting and occlusions. Although long-term tracking and detailed telemetry are unavailable, this approach enables SMEs to achieve key insights into demographic segmentation and purchasing patterns, akin to e-commerce capabilities. Compliance with operational and legal standards, including POPIA and GDPR, is maintained. This study contributes to the literature by presenting a comprehensive model that integrates multiple technologies into a single pipeline for SMEs, while accounting for scalability and privacy. Future recommendations include phased deployment and scaling considerations, forming a foundation for SMEs to enhance their retail analytics and compete more effectively in developing markets..

Table of Contents

Chapter 1 — Introduction 1
1.1 Background and Motivation 1
1.2 Problem Definition 1
1.3 Research Objectives 1
1.4 Research Questions 2
1.5 Feasibility Focus for SMEs 2
1.6 Delivering E-Commerce-Level Insights for Physical Retail 2
1.7 Structure of the Thesis 2
Chapter 2 — Literature Review 3
2.1 CCTV-Based Analytics in Retail 3
2.2 Computer Vision for Demographics 4
2.3 Face Detection and Age/Gender Estimation 5
2.4 YOLO Architectures 6
2.5 DeepFace and Facial Embeddings 8
2.6 Retail Analytics: E-Commerce vs In-Store 8
2.7 SMEs and Digital Capability Gaps 9
2.8 Ethical and Legal Considerations (POPIA/GDPR) 11
Chapter 3 — Research Methodology 12
3.1 Research Design 12
3.2 Data Collection Strategy 12
3.3 Sampling Period and SME Constraints 13
3.4 Technical Pipeline 14
3.4.1 Person Detection (YOLO) 14
3.4.2 Face Cropping 15
3.4.3 Facial Embeddings & Identity Grouping 16
3.4.4 Age & Gender Prediction 17
3.5 Evaluation Metrics 18
3.6 Limitations and Feasibility of Cloud Vision Analytics Infrastructure 19
Chapter 4 — Research Findings 20
4.1 Prediction Accuracy Results 20
4.2 Prediction Accuracy Results 22
4.3 Demographic Distribution Results 26
4.4 Revenue & Transaction-Based Analytics Output 29
4.5 Temporal Patterns in Revenue and Transactions 32
Chapter 5 — Discussion 34
5.1 Alignment with Literature 34
5.2 Feasibility for SMEs 36
5.3 Comparison to E-Commerce Analytics Benchmarks 37
5.4 Data Quality & Model Performance Analysis 38
5.5 Insights for Retail Marketing Strategy 40
5.6 Realistic Operational Considerations for SMEs 41
Chapter 6 — Conclusion 43
6.1 Summary of Research 43
6.2 Final Evaluation of SME Feasibility 43
6.3 Contributions of This Study 44
6.4 Recommendations for Implementation 44
6.5 Limitations 45
6.6 Future Research 45

List of Figures

Figure 1 : Gender Confidence Distribution 23
Figure 2 : Gender Confidence by Predicted Gender 23
Figure 3 : Gender Confusion Matrix 24
Figure 4 : Age Confidence Distribution 24
Figure 5 : Age Confidence by Predicted Gender 25
Figure 6 : Age Confidence by Predicted Age 25
Figure 7 : Age Confusion Matrix 26
Figure 8 : Transaction Amount by Gender 27
Figure 9 : Transaction Amount by Age Group 27
Figure 10 : Revenue by Gender 28
Figure 11 : Revenue by Age 28
Figure 12 : Heatmap of Revenue by Gender and Age 29
Figure 13 : Category Revenue by Gender 30
Figure 14 : Category Purchases by Gender 30
Figure 15 : Top 10 Items by Age Group 31
Figure 16 : Count Line Items by Age Group 31
Figure 17 : Total Revenue by Age Group 32
Figure 18 : Hourly Transaction Pattern by Gender 33
Figure 19 : Weekday Transaction by Age Group 33
Figure 20 : Hourly Transactions by Age Group 34
Figure 21 : Hourly Revenue by Age Group 34
Figure 22 : Hexbin Joint Age and Gender Confidence Distribution 59

List of Tables

Table 1 : Summary of Person Detection Parameters 15
Table 2 : Key Face Cropping Parameters 16
Table 3 : Facial Embeddings and Clustering Parameters 17
Table 4 : Summary of Age and Gender Prediction Models 18
Table 5 : Demographic Accuracy Metrics 19
Table 6 : Business-Focused Evaluation Categories 19
Table 7 : Estimated Cloud Vision Analytics Infrastructure Costs and Requirements 20
Table 8 : Comparison of Age and Gender Prediction Performance Between prithivMLmods and nateraw Models 22

List of Abbreviations

Abbreviation	Full Term
AI	Artificial Intelligence
AWS	Amazon Web Services
CCPA	California Consumer Privacy Act
CCTV	Closed-Circuit Television
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CSP	Cross Stage Partial
CSV	Comma-Separated Values (data format)
DLDL	Deep Label Distribution Learning
DNN	Deep Neural Network
DPIA	Data Protection Impact Assessment
EC2	Elastic Compute Cloud
FSRCNN	Fast Super-Resolution Convolutional Neural Network
GDPR	General Data Protection Regulation
GELAN	Gradient-Enhanced Layer Aggregation Network
GFPGAN	Generative Facial Prior Generative Adversarial Network
GPU	Graphics Processing Unit
HCI	Human-Computer Interaction
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
HOG	Histograms of Oriented Gradients
ICO	Information Commissioner’s Office
IoT	Internet of Things
IT	Information Technology
LFW	Labeled Faces in the Wild
MAE	Mean Absolute Error
mAP	Mean Average Precision
MTCNN	Multi-task Cascaded Convolutional Neural Network
NMS	Non-Maximum Suppression
NVR	Network Video Recorder
ODBC	Open Database Connectivity
OECD	Organisation for Economic Co-operation and Development
PAN	Path Aggregation Network
PGI	Programmable Gradient Information

Abbreviation	Full Term
POE	Power over Ethernet
POPIA	Protection of Personal Information Act
POS	Point-of-Sale
R-ELAN	Residual Efficient Layer Aggregation Network
RFID	Radio-Frequency Identification
SME	Small and Medium-Sized Enterprise
SSD	Solid State Drive
SVM	Support Vector Machine
UAV	Unmanned Aerial Vehicle
UMAP	Uniform Manifold Approximation and Projection
ViT	Vision Transformer
YTF	YouTube Faces Dataset
YOLO	You Only Look Once (object detection model)

Chapter 1 — Introduction

1.1 Background and Motivation

Physical retail SMEs struggle to match the data-driven sophistication of online retailers, who leverage detailed user analytics to optimise strategy. Traditional retailers often find themselves in the dark; a local shop owner, for instance, might guess which items are popular based on sporadic customer feedback, whereas an online rival can pinpoint exact purchase patterns through tracking every click. This reliance on limited sales data and manual observation hinders understanding of customer behaviour. Advances in computer vision and the increased use of CCTV now allow for automated, affordable in-store analytics, such as quantifying foot traffic and profiling demographics. This thesis investigates how these technologies can offer SME retailers valuable marketing intelligence once limited to larger enterprises.

1.2 Problem Definition

Despite the availability of open-source computer vision tools, SMEs face challenges such as limited hardware resources, insufficient technical skills, and privacy compliance concerns. These barriers restrict their use of in-store analytics to inform marketing and operations. If SMEs delay adopting such analytics, they risk losing competitive ground, potentially leading to an estimated 15% decrease in revenue over the next five years due to missed opportunities in targeted marketing and customer engagement. Therefore, there is a clear need for an accessible, practical system that demonstrates how meaningful customer insights can be produced for SMEs under real-world constraints.

1.3 Research Objectives

This survey investigates whether CCTV footage can provide adequate demographic insights in small- and medium-sized retail businesses. The intent is to develop a comprehensive analytical pipeline that detects people in video, analyses faces, and visualises the results using publicly available open-source tools. In addition, models including YOLO for person detection, DeepFace for face analysis, and selected age- and gender-related predictors from Hugging Face will be used to create customer data comparable to that used by online retailers. The research also investigates the technical, legal, and practical implications of such a system in actual SME settings.

1.4 Research Questions

To achieve these goals, this study explores several vital research questions. First, it examines the accuracy of open-source models to estimate demographic properties, in this case, age and gender, from CCTV data in retail environments. Second, it explores which in-store analytics can be derived from such data to optimise decision-making in SMEs. Third, the study investigates whether the proposed analytical system is efficient under typical SME constraints, including limited computational resources and budgetary restrictions. Lastly, it assesses how the insights developed by this approach compare with the analytics capabilities commonly available on e-commerce platforms.

1.5 Feasibility Focus for SMEs

Instead of having specialised analytics teams and proprietary tools, as big retailers do, SMEs typically have to run their businesses with a relatively tight budget, basic hardware, and little or no technical know-how. In this paper, we directly address these challenges by leveraging existing CCTV systems, freely available accessible models, and modest datasets. This will determine whether meaningful customer insights can be obtained without high cost and tech investment, and whether it would be feasible, especially given the normal operations SME employees face.

1.6 Delivering E-Commerce-Level Insights for Physical Retail

E-commerce companies benefit from detailed behavioral data, enabling precise targeting and operational optimization. Physical retailers, by contrast, lack comparable insight into in-store consumer actions and pathways. This work directly addresses whether computer vision-based analytics—using CCTV to extract demographic profiles, visit frequencies, and peak hours—can provide SMEs with actionable, e-commerce-level marketing insights. The central argument is that if affordable in-store computer vision tools can supply results similar to online analytics, SMEs can close the data advantage gap and compete more effectively in modern retail..

1.7 Structure of the Thesis

The outline of this thesis is presented in six chapters. Chapter 1 presents the research context, motivation, objectives, and the main research questions that will define the study. Chapter 2 is an overview of the literature on topics such as computer vision, facial analytics, SME digital transformation, and the ethical and legal frameworks. Chapter 3 discusses the research methodology, outlining in detail the data collection, model selection, implementation

of the analytical pipeline, and the evaluation metrics used. Chapter 4 presents empirical findings, including demographic determinations, visual analytics outputs, and the consolidation/integration of combined datasets. Chapter 5 examines the interpretation of the findings within the literature, the feasibility of the study for SMEs, and the broader implications of the study. Then Chapter 6 wraps up this thesis by summarising contributions, outlining limitations, and proposing recommendations for future research.

Chapter 2 — Literature Review

2.1 CCTV-Based Analytics in Retail

CCTV entered retail in the 1970s as basic black-and-white security cameras for detecting theft (ThinkLP, 2023). The 1980s brought colour imaging, motion detection, and cheaper recorders, expanding its use to broader store and mall monitoring (DTiQ, 2023). In the 1990s, digitisation led by DVRs enabled greater video quality and remote control over content as CCTV became mainstream (Connell, Fan, Gabbur, Haas, & Pankanti, 2013)..

Subsequent to the 2000s, high-definition IP cameras, the development of wireless networking, and advances in artificial intelligence and computer vision transformed CCTV technology from a passive security device into an embedded analytics tool that provides real-time behavioural insights (Connell et al., 2013; Pelco by Motorola Solutions, 2023). Previous systems were largely focused on manual loss prevention (Senior et al., 2007), whereas more recent smart-video solutions automate business intelligence measurement, leading to KPIs such as footfall, queue length, and shopper engagement (Connell et al., 2013; Tictag, 2023).

Common applications include:

- Footfall and Traffic Tracking: Automation of counting based on visitor counts, volume, and group sizes achieved an accuracy of about 85–95% in a controlled environment (Senior et al., 2007; Tictag, 2023).
- Heat Maps and Dwell Time: Overhead cameras reveal where staff are focused, providing cues for merchandising decisions (Senior et al., 2007; Tictag, 2023).
- Real-time queue monitoring helps avoid crowds and improve checkout flow, a widely used feature in larger retail chains (Pelco by Motorola Solutions, 2023).
- Shopper flow and behavioural analysis: Trajectory tracking reveals bottlenecks and behavioural trends associated with product engagement (Connell et al., 2013; Tictag, 2023).

- Advanced behavioural insights: AI facial analytics can provide insights on anonymised age, gender, and responses to market displays (Pelco by Motorola Solutions, 2023; Tictag, 2023).

Benefits reported as being attributed to enhanced sales, including sales uplift by 3–5%, 2–3% reduction of operating expenses, superior customer satisfaction, enhanced loss-prevention measures (Beck, 2020; Pelco by Motorola Solutions, 2023; Tictag, 2023). Continued barriers to adoption include occlusions, lighting variation, infrastructure complexity, regulatory obligations (GDPR/CCPA), and high upfront investment, which remain significant barriers for SMEs (Beck, 2024; Growth Market Reports, 2025; Hodgson, 2025).

While large retail groups continue to lead the adoption trend, smaller firms have seen scalable cloud-based analytics solutions significantly reduce costs and knowledge/experience thresholds (Growth Market Reports, 2025; Hodgson, 2025; ThinkLP, 2023). Used ethically, modern CCTV analytics represent a disruptive technology that leverages security and strategic data insights to increase profits and customer satisfaction (Beck, 2024; Connell et al., 2013).

2.2 Computer Vision for Demographics

Computer vision models for demographic estimation, based on approaches such as age and gender, are now used across various research and industry domains (Buolamwini & Gebru, 2018; Rothe, Timofte, & Van Gool, 2015). Soft biometric cues from facial images enable distinguishing people without uniquely identifying them (Jain et al., 2010). Although gender and age are prominent social cues, age estimation remains challenging as perceived and chronological age differ substantially (Guo & Mu, 2011; Rothe et al., 2015). Androgynous, culturally different and heterogeneous facial features complicate the classification of gender (Levi & Hassner, 2015).

Earlier approaches utilised handcrafted (wrinkles, texture, geometric ratios) signals and employed classical classifiers such as SVMs or Random Forests, but performed poorly under changes in lighting, stance, and expression (Guo & Mu, 2011; Khan et al., 2020). Deep learning, and in particular CNNs, surpassed these designs by learning features from pixels (Rothe et al., 2015; Swaminathan et al., 2020). Significantly, in its DEX model, derived from VGG-16 and trained on IMDB-WIKI, a state-of-the-art accuracy was achieved with MORPH and FG-NET (Rothe et al., 2015). Transfer learning with FaceNet also works well for gender and ethnicity classification (Swaminathan et al., 2020).

More recently, advanced techniques comprise multi-task networks, region-based models, and mask- and occlusion-resilient approaches (Alghaili, Li, & Ali, 2020; Khan et al., 2020). Transformer-based architectures further improve performance by modelling global context via self-attention over the facial area (Shi et al., 2023).

Even so, accuracy across demographic groups is not uniform, as training biases and dataset imbalance contribute to this (Buolamwini & Gebru, 2018; Klare et al., 2012). Error rates tend to be higher among darker-skinned women and older adults (Buolamwini & Gebru, 2018; Raji et al., 2020). Mitigation includes balanced datasets (FairFace and UTKFace) and better training regimes (Karkkainen & Joo, 2021).

These advances make demographic inference attractive for retail analytics, targeted marketing, and interactive systems, but fairness, privacy, and consent must be emphasised to prevent misuse.

2.3 Face Detection and Age/Gender Estimation

Face detection serves as the foundation for facial analytics, enabling us to make downstream predictions about demographic behaviour. Early detectors like Viola–Jones's Haar Cascades, which delivered fast frontal-face detection, were very sensitive to pose, occlusions, and lighting issues (Viola & Jones, 2001; OpenCV, 2022). The HOG–SVM algorithms improved robustness to illumination, though they still failed to address non-frontal views and small faces (Dalal & Triggs, 2005; King, 2009).

Deep learning models achieved significant improvements in detection thanks to direct feature learning. Architectures that combine detection and landmark localisation, such as MTCNN, have strong multi-scale capability, albeit at an increased computational cost (Zhang, Zhang, Li, & Qiao, 2016). SSD and YOLO-based single-shot detectors have been shown to achieve near real-time performance on face detection (Liu et al., 2016; Bochkovskiy, Wang, & Liao, 2020), and RetinaFace further enhances accuracy with multi-task learning and 3D pose estimation (Deng, Guo, Niannan, & Zafeiriou, 2020). Lightweight models like YuNet achieve faster inference and better occlusion processing than classical cascades, albeit at lower accuracy than heavier CNN models (Yu, Qi, & others, 2023; OpenCV, 2022). In practice, systems combined detectors for a speed-precision trade-off (Serengil & Ozpinar, 2020).

Age estimation methods differ widely. The regression-based models treat age as a continuous variable but are sensitive to noisy labels and have difficulty with extreme ages (Lanitis, Taylor, & Cootes, 2002; Ranjan et al., 2016). Classification models predict single-year or bracketed classes (Rothe et al., 2015), while Label Distribution Learning (DLDL) spreads probability across neighbouring bins to increase the robustness of those classes; this has also proven advantageous for large datasets (Gao et al., 2018). Such ordinal regression formulations leverage the ordering of age labels, as noted by Liu et al. Recent work has demonstrated that soft-label distributions and ordinal methods are more relevant for representing the uncertainty in perceived age (Agbo-Ajala et al., 2022).

Gender classification is generally considered a binary task, and CNN models can achieve>95% accuracy on balanced datasets (Levi & Hassner, 2015; Karkkainen & Joo, 2021). A multitask-based architecture for facial recognition and attribute prediction achieves very high accuracy in controlled experiments (Ranjan et al., 2016).

But even this demographic modelling has its fairness issues ahead. For instance, the Gender Shades study reported large error rates for darker-skinned females than for light-skinned males (Buolamwini & Gebru, 2018). Balanced datasets, such as FairFace, reduce such inequalities by enhancing representation (Karkkainen & Joo, 2021).

Well-annotated datasets depend heavily on the model's overall reliability. IMDB-WIKI remains the largest publicly available age/gender database, with over 500,000 labelled images (Rothe et al., 2015). UTKFace (Zhang, Song, & Qi, 2017), Adience (Eidinger, Enbar, & Hassner, 2014), and FairFace (Karkkainen & Joo, 2021) offer varying levels of diversity and challenging environments. Pre-trained models, including DEX, or lightweight models such as CNNs from Hugging Face, can be deployed quickly with good accuracy—typically MAE of 2.5–3 years for age and >95% for gender (Rothe et al., 2015; fanclan, 2023; Agbo-Ajala et al., 2022).

2.4 YOLO Architectures

The YOLO family is a fast-evolving architecture, and currently, versions 8 through 12 are among the best real-time object detectors. YOLOv8 proposes an anchor-free design with a decoupled detection head and achieves better accuracy—especially for small and overlapping objects—while retaining strong speed characteristics. Its backbone features the C2f (CSP-Fast) module and the PAN neck with dynamic kernel attention, and the system is

used for detection, segmentation, pose estimation, and classification (Jocher, Chaurasia, Qiu, & Ultralytics, 2023).

YOLOv9 took this further with reversible residual layers based on the Information Bottleneck Principle and a programmable gradient mechanism (PGI) that enables deeper feature learning. GELAN feature aggregation further enhances mAP (Sapkota et al., 2025), but introduces slightly higher inference latency, indicative of the continuing trade-off between accuracy and speed.

YOLOv10 was a step forward on end-to-end detection by removing NMS altogether. A one-to-many training head and a one-to-one inference head make distinctive detections with lower computational load. The solution architecture comprises state-of-the-art CSP modules, rank-guided pruning, and large-kernel convolutional blocks, enabling lighter models to be used more effectively on edge devices (Sapkota et al., 2025).

YOLOv11 mainly focused on ultra-efficient multi-tasking. It substitutes the C2f blocks of YOLOv8 with C3k2-style bottlenecks and introduces the C2PSA module to improve small-object performance via spatial self-attention. It retains YOLOv10's NMS-free feature and offers the fastest inference speed in the series—perfect for real-time applications such as live monitoring and UAV systems (Jocher, Qiu, & Ultralytics, 2024).

YOLOv12 employs methods from the transformer-inspired architecture while maintaining real-time performance. Its Area Attention (A^2) module abstracts larger receptive fields by regionally partitioning feature maps, improving robustness in cluttered scenes. YOLOv12 also utilises an R-ELAN model to stabilise training with attention layers, while using FlashAttention for high-speed transformer computation. The highest mAP in COCO is approximately 55.2%, which can be attributed to the successful convergence between convolutional and attention-based designs (Tian et al., 2025).

Overall, YOLOv8–v12 demonstrates clear technical advancement, from anchor-free prediction and decoupled heads to NMS-free detection and transformer-augmented attention, positioning the YOLO family as cutting-edge for retail analytics, surveillance, and any application demanding prompt, precise detection.

2.5 DeepFace and Facial Embeddings

Face recognition has replaced traditional image-matching methods with embedding-based techniques, in which deep neural networks map facial images to compact feature vectors that cluster by identity and separate individuals (Schroff, Kalenichenko, & Philbin, 2015; Taigman, Yang, Ranzato, & Wolf, 2014). In this sense, those embeddings enable verification, as well as identification and clustering via similarity metrics such as cosine distance.

DeepFace, launched by Facebook AI Research, was among the first systems to achieve human-level verification accuracy. It used a nine-layer CNN, 3D frontalization for alignment, and 4096-dimensional embeddings, achieving 97.35% accuracy on LFW (Taigman et al., 2014). FaceNet performed better when trained with triplet loss, resulting in compact 128-dimensional representations and 99.63% accuracy on LFW (Schroff et al., 2015). VGG-Face was also designed and trained from a deeper VGG-16 architecture trained on 2.6 million images, achieving similarly strong results over 4096-dimensional embeddings and high accuracy on LFW and YTF (Parkhi, Vedaldi, & Zisserman, 2015).

ArcFace pushed the field further with an additive angular-margin loss, resulting in a more pronounced angular separation between identities. ResNet-100 backbone, normalised 512-dimensional embeddings, achieved over 99.8% accuracy on LFW and were generally robust to pose and age variation (Deng et al., 2019).

The DeepFace Python library (Serengil & Ozpinar, 2020) aggregates these (DeepFace, FaceNet, VGG-Face, ArcFace) and creates a reusable pipeline for detection, alignment, embedding extraction, and optional attribute prediction (age, gender, emotion, and ethnicity).

The fundamental operation of these systems is the same: face recognition is performed in a single embedding space, intended to achieve both intra-class similarity and inter-class separation (Deng et al. 2019; Schroff et al. 2015). These challenges are ongoing, including demographic fairness, robustness to occlusions and extreme poses, and privacy-compatible deployment (Kotwal & Marcel, 2024; NIST, 2019).

2.6 Retail Analytics: E-Commerce vs In-Store

E-commerce has previously led in retail analytics because every touchpoint between users (views, clicks, add-to-cart events, and purchases) is strictly digital. This also allows you to track a behavioural trend or a personalised product recommendation in real time, track it with

A/B testing precision, and strongly link the customer journey to your advertising campaign with marketing attribution (IoT For All, Renno, 2024; RudderStack, Varangaonkar, 2021).

Conventional physical stores, by contrast, have had far less visibility, primarily using aggregated metrics such as footfall or total sales. This gap in “analytics” for in-store behaviour, dwell times, and how well merchandising was performing made these largely invisible without specialised hardware, which has traditionally been costly and complicated for many retailers—particularly SMEs—to implement (Zensors, 2023; RetailNext, Kirsten, 2023).

Recent breakthroughs are bridging this gap. IoT sensors, AI-based computer vision, RFID, and integrated omnichannel data sets are enabling physical stores to collect insights that were once the preserve of e-commerce. IoT beacons can connect in-store actions to loyalty profiles, and computer vision generates live heatmaps, product-interaction metrics, and demographic insights (IoT For All, Renno, 2024; Sakovich, 2020). Amazon Go shows how these methods are combined to create fully automated shops, where every customer action is recorded in real time (Lumenalta, 2025). The use of RFID by the likes of Adidas and Levi’s provides evidence that real-time stock visibility can minimise the need for such manual inventory management (LS Retail, Jonnson, 2024). In addition to online browsing and purchasing history, retailers have a unified omnichannel view for all their marketing attribution and inventory management (Admetrics, 2025).

There are incremental barriers to entry for SMEs (small and medium-sized businesses) that cloud-based tools gradually eliminate, including hardware costs, technical skills, and operational upkeep (Business.com, Farlie, 2025). Retailers using analytics report improvements in staffing, conversion, and organisational performance, driven by greater efficiency (RetailNext, Kirsten, 2023).

As a result, although e-commerce used to reign supreme in retail analytics, thanks to its digital infrastructure, physical retail is quickly catching up. We see the future of hybrid retail models that blend online and in-store data to provide consumers with tailored, effective, and extremely agile shopping solutions (MIT Sloan Management Review, 2025).

2.7 SMEs and Digital Capability Gaps

The biggest gaps in SMEs’ digital capabilities are well established and have been constraining meaningful digital transformation. A large part of this starts with infrastructure:

some have outdated systems, unreliable connectivity, and data silos that render them less competent at accessing, storing, and integrating data (Enefer, 2023; Patterson-Waites, 2023). These issues are particularly prominent in the developing world as they contribute to wider digital divides (OECD, 2021). SMEs often lack the technical capability to employ modern data analysis or automation tools (Kgakatsi et al., 2024) due to the absence of dedicated IT teams and scalable platforms provided by large enterprises.

Skills shortages contribute to another obstacle. SMEs face the challenge of gaining expertise in data science, analytics, and cybersecurity and are often confronted with an organisational culture that lacks data-driven practices to enable adoption (Enefer, 2023; Long, 2025). The evidence from the study suggests that, in the presence of digital tools, SMEs frequently underutilise data or derive little value from investments, due to low data literacy levels and poor management support (World Economic Forum & Bainchini, 2019).

There are also financial problems, which inhibit progress. Tight budgets are also limiting for investment in advanced technology or trained staff, and continued subscription or maintenance costs add pressure (Long, 2025; Patterson-Waites, 2023). Small teams also mean staff have to do many tasks in a small firm, leaving little time to pilot or maintain complex digital projects (Enefer, 2023). Therefore, analytics and automation feel more like a gamble than a need. For this, pay-as-you-go cloud models or financial incentives remain among the recommended measures (Long, 2025; OECD, 2021).

These constraints exacerbate a wider competitive imbalance: big companies have access to capital, specialist analytics staff, and well-functioning digital infrastructures, which enable broad uptake of big data and AI. In contrast, only 10% of small European companies use big data analytics, compared to 33% of large firms (OECD, 2021). This “liability of smallness” perpetuates the disadvantage faced by SMEs (Patterson-Waites, 2023). In such a setting, the agility of SMEs, together with low-code tools and cloud services, provides a path to close the gap (Enefer, 2023; Long, 2025).

Closing this capability gap will require both internal development – in terms of skills, data literacy, and culture – and external assistance, including access to technology, training, and funding. Combined, these initiatives enhance SMEs’ capacity to compete, innovate, and build resilience in a rapidly growing digital economy (Long, 2025; OECD, 2021; World Economic Forum & Bainchini, 2019).

2.8 Ethical and Legal Considerations (POPIA/GDPR)

Image-based technologies for demographic analysis in retail pose considerable ethical and legal obligations under GDPR in the EU and POPIA in South Africa. Both biometric and facial data are considered sensitive personal information and, therefore, are subject to stringent rules of conduct regarding transparency, consent, data minimisation, purpose limitation, and individual rights.

Under GDPR, facial images and biometric identifiers are regarded as special categories of data (GDPR 2016, Arts. 4[14], 6, 9). Typically, retailers must obtain informed and unambiguous consent — a challenge in public retail settings where passive observation renders explicit consent virtually impossible (ICO, 2022; Lewis Silkin LLP, Laher, 2025). “Legitimate interest” (Art. 6(1)(f)) potentially could be operationalised theoretically; a rigorous necessity and proportionality analysis is required to process it, and it by itself does not apply for biometric data without some further Article 9 exception (ICO 2022; Reuters 2020).

POPIA also imposes similar tight requirements. Biometric information is considered special personal information and processing without express consent or legislative authority is prohibited (POPIA, 2013; Michalsons, 2024). POPIA contains requirements for purpose specification, openness, strict security, and limits on retention; the Information Regulator (Data Protection Laws of the World, 2025), for example, is responsible for both the design and administration of protection measures.

The ethical concerns are not limited to compliance. Many facial recognition systems exhibit demographic bias, leading to disproportionate error rates for individuals from underrepresented groups and increasing the risk of fairness bias (Keylabs, 2024; Xiang, 2022). Responsible deployment thus requires the use of a variety of datasets, checks for bias, and adherence to principles of fairness and non-discrimination (ICO, 2022; Keylabs, 2024). Wider risks include excessive surveillance, profiling, and chilling effects on customer behaviour, thus justifying the need for clear information channels, opt-in controls, and tight purpose limitation (Lewis Silkin LLP, Laher, 2025; Reuters, 2020; Tencent Cloud, 2024).

GDPR and POPIA mandate secure storage, limited retention, and timely deletion, along with significant security features such as encryption, pseudonymization, and restricted access (GDPR, 2016; POPIA, 2013; Michalsons, 2024). Before deployment, Data Protection Impact Assessments (DPIAs) are also key to assessing risks and exploring less intrusive alternatives (ICO, 2022; Lewis Silkin LLP, Laher, 2025).

Retailers using computer vision for demographic analytics must operate within strict regulatory frameworks while upholding ethical principles of fairness, transparency, and respect for autonomy—principles that are key to maintaining trust in increasingly monitored retail environments (Lewis Silkin LLP, Laher, 2025; Xiang, 2022).

Chapter 3 — Research Methodology

3.1 Research Design

This study adopts an applied, experimental case study design to determine if it is feasible to implement computer vision analytics using existing CCTV infrastructure in a small-to-medium enterprise (SME) retail setting. This mixed-methods approach combines quantitative analysis of model outputs—age, gender, and visit frequency—with qualitative feedback from customers and the store owner to assess both the system's accuracy and its perceived business value.

This is the basis for developing a computer vision pipeline using open-source software tools deployed on low-cost hardware, while accounting for SMEs' resource constraints. Based on the researcher’s experience in a family-owned retail business, the project aims to leverage existing security-focused CCTV deployments to generate marketing and customer insights comparable to those typically offered on e-commerce platforms.

3.2 Data Collection Strategy

Real-world CCTV footage and point-of-sale (POS) transaction data from an operational café were used to guarantee ecological validity and practical relevance in this study. One UniFi Protect camera was mounted statically on a shelf above the checkout area—where customers place and pay for their orders—offering a high-engagement vantage point to observe key facial attributes and in-store interactions. Under natural customer traffic, the camera recorded 1920 × 1080-pixel video continuously for two full business days (Ubiquiti Inc., 2023a).

To accelerate scalable, automated data recovery, footage was downloaded from 08:00 to 17:00 in hourly batches using the uiprotect Python API, an unofficial interface for UniFi Protect systems (uiprotect, 2024). This enabled automated extraction directly from the store’s local network video recorder (NVR), which effectively imaged a real-world integration

pathway for SMEs with comparable surveillance facilities. Audio recordings of the data were not included in the analysis, since the data were collected only by the camera.

Concurrently, structured transaction records were generated from the café's SQL Server–based POS system through an authenticated ODBC connection. The day data set was filtered, and the result was detailed itemised receipts with timestamps, sales categories and quantities. This multimodal data enabled cross-validation of computer vision-derived metrics, including footfall and demographic predictions, against actual sales data, deepening the feasibility evaluation through integrated analysis.

Ethical implications and privacy procedures were considered during data collection. The video footage was already there, acquired as part of the café's regular security perimeter. Video data was processed as needed for both security and analytics, and customers were informed using on-site signage. All videos were securely deleted after processing, with only anonymised facial embeddings retained for downstream analysis. No personal or biometric information was kept, and no users' data was saved.

Although the implementation used UniFi Protect and SQL Server, the data collection approach can be extended to other retail environments with image cameras and POS databases that support data export. The multimodal, privacy-conscious model we developed is an example of how SMEs utilise their infrastructure to generate cost-effective operational insights.

3.3 Sampling Period and SME Constraints

This study used real-world footage collected from a working café over two full business days, Wednesday, November 5th, and Friday, November 7th, 2025. Recording sessions took place from 08:00 to 17:00 every day to gather both mid-week and weekend-preparatory customer activity. A static UniFi Protect G4 camera was mounted above the point-of-sale (POS) counter, where many customers place and complete their orders, yielding the best vantage point for facial visibility and behavioural data capture. Thus, 18 hours of unedited, natural working conditions footage for video was made from the footage.

The camera recorded a 1920×1080 video, captured images, and saved them to the café network video recorder (NVR) and the local machine. Roughly 135 GB of raw video files were downloaded through the unofficial uiprotect Python API and saved to the researcher's local solid-state drive (SSD) for batch processing. Sound had been recorded, and audio was

excluded from the analytical workflow. Raw videos were deleted, and only anonymised facial embeddings were saved for further analysis.

This setup provides a more realistic baseline for SMEs, which currently use CCTV systems primarily for security. The café owns the majority of the infrastructure, such as camera hardware, cabling, network infrastructure (via UniFi Dream Machine Pro), and a centralised NVR, which effectively lowered the barriers to integration.

Locally, data was processed and evaluated using the researcher's personal workstation. The full system specifications used for all experiments are provided in Appendix B (Table B1: System Specifications), ensuring transparency and reproducibility while keeping technical details out of the main body of the thesis.

Though local infrastructure is sufficient for short-term analysis, scaling sampling time to the volumetric scales common for long-term data acquisition (e.g., a full month) would likely exceed local processing and storage capacity. Scalability would require the use of cloud infrastructure, featuring storage solutions, GPU-accelerated virtual machines, and automated data pipelines.

In addition to video data, transactional information was accessed from the café's cloud-hosted POS system to compare footfall with actual purchasing behaviours. Such multimodal integration allows for deeper analysis of flow characteristics and buying behaviours.

3.4 Technical Pipeline

3.4.1 Person Detection (YOLO)

The pipeline begins with person detection using the YOLOv11s model (Ultralytics, 2024), chosen for its speed and efficiency on mid-range GPU hardware. Footage was captured from a fixed 1080p overhead camera positioned above the POS counter, and synchronised with POS transaction timestamps from the accounting export. Only frames that occurred shortly after each transaction—during moments when customers typically face the counter—were evaluated.

YOLOv11s was applied to identify class ID 0 (person) and produce the candidate regions that would be carried forward to the face-processing stages. Once detections were registered, the corresponding frame, timestamp and transaction metadata were saved, and a secondary pass isolated individual person boxes to prepare clean inputs for subsequent facial analysis.

Video handling was performed asynchronously via FileVideoStream to reduce I/O bottlenecks during frame extraction (Rosebrock, 2024). Each frame was matched to the closest transaction time to prevent duplicate saves and maintain a clean 1-to-1 mapping. The entire workflow was implemented using the Ultralytics Python library and documented in notebook 3_yolo_person.ipynb (Wolf, 2025i).

All identifiable frames were temporary and deleted after processing, with only anonymised embeddings retained to ensure compliance with applicable privacy regulations.

Parameter	Setting	Purpose
Confidence Threshold	≥ 0.8	Retain detections with high confidence
Bounding Box Height	≥ 30% of frame height	Ensure detection represents close-up person
Bounding Box Width	≥ 20% of frame width	Ensure detection represents close-up person
Cropping Threshold	≥ 80 × 80 pixels	Generate isolated person images for analysis
Processing Time Window	~1.75 minutes post-transaction	Focus on customer-facing moments
FPS	25 (assumed)	Calculate frame timestamps

Table 1: Summary of Person Detection Parameters

3.4.2 Face Cropping

After the person-level detections were completed, a two-stage face-cropping process was used to isolate and refine facial images for later analysis. DeepFace was employed with the YuNet backend, which proved reliable on CCTV-quality footage and handled the variability of angles and lighting reasonably well (Yu, Qi, & others, 2023; Wolf, 2025g).

Each detected person image from the previous stage was passed through this detector, and a light round of pre-processing was applied to stabilise contrast and improve clarity before extraction. This ensured that the facial regions remained usable once aligned with the transaction timestamps.

A secondary enhancement and filtering step followed. Super-resolution was tested as an optional improvement step, and although FSRCNN produced visibly sharper crops, GFPGAN was removed from the final pipeline due to its inconsistent behaviour under the café's lighting

conditions. A final frontal-face filter was applied to exclude profiles or misaligned detections, keeping only clean forward-facing images for the age and gender models. The complete implementation is documented in 5_image_processing.ipynb (Wolf, 2025h).

Parameter	Setting	Purpose
Face Detector	YuNet (via DeepFace)	Efficient face detection on CCTV frames
Detection Confidence	≥ 0.65	Filter unreliable detections
Crop Size	192–384 pixels (resized)	Standardize input dimensions
Expansion Padding Ratio	0.5	Include contextual facial details
Preprocessing Techniques	CLAHE, Gamma correction	Enhance contrast and tone balance
Super-resolution Model	OpenCV FSRCNN (4× upscale)	Improve image clarity
Face Restoration Model	GFPGAN (disabled final)	Attempted face restoration (inconsistent)
Frontal Face Filtering	Haar cascade (frontalface_alt2)	Exclude profiles/misdetections

Table 2: Key Face Cropping Parameters

3.4.3 Facial Embeddings & Identity Grouping

After producing clean frontal face crops, the next stage involved generating facial embeddings and grouping recurring customers into identity clusters. This allowed each customer to be reliably linked back to their POS record for downstream demographic analysis.

Embeddings were produced using the Facenet512 model in the DeepFace framework, which demonstrated stable performance on CCTV-quality imagery (Schroff et al., 2015; Serengil & Ozpinar, 2024). YuNet was again used as the detection backend to confirm face presence in every crop (Yu, Qi, & others, 2023), and each embedding was stored together with its associated metadata for later merging (Wolf, 2025c).

The embeddings were normalised and then clustered using HDBSCAN, which handled the varying density and noise patterns common in CCTV footage (HDBSCAN, 2016). Noise points were reassigned where appropriate, and a secondary pass with more relaxed settings was used to capture smaller identity groups that might otherwise be treated as outliers.

To inspect cluster structure, the embeddings were also projected into two dimensions using UMAP, which gave a clearer picture of how identities separated or overlapped in practice (UMAP, 2024). When clusters represented the same individual, merging was performed using cosine-based similarity measures to avoid splitting a single customer into multiple identities.

Each identity cluster was then matched back to transaction data through filename-encoded timestamps and AccoID fields, and within each cluster, the most frontal representative image was selected for demographic inference. This step finalised the link between faces and transaction records, enabling the demographic analytics used in later chapters (Wolf, 2025d).

Because embeddings scale into the thousands, the clustering process is sensitive to dataset size and quality, and filename consistency proved essential for accurate mapping. Throughout the workflow, all files containing identifiable information were handled in line with data protection requirements.

Step	Parameter	Value/Setting	Purpose
Face Embedding Model	Facenet512	High accuracy model	Generate discriminative facial vectors
Face Detection Backend	YuNet	Efficient face detection	Enforce face presence per image
Embedding Normalization	sklearn.preprocessing.normalize	Unit length normalization	Prepare embeddings for cosine similarity
Clustering Algorithm	HDBSCAN	min_cluster_size=6, min_samples=2	Density-based clustering with noise detection
Noise Reassignment Threshold	Cosine distance	< 0.25	Assign noise to closest cluster centroid
Reclustering	HDBSCAN	Relaxed parameters	Improve coverage with micro-clusters
Dimensionality Reduction	UMAP	n_neighbors=15, min_dist=0.1, metric=cosine	Visualization of embedding clusters
Cluster Merging Threshold	Cosine similarity	≥ 0.85	Merge similar, small clusters
Representative Image Selection	Haar cascade heuristic	Aspect ratio and centering score	Select most frontal image per cluster

Table 3: Facial Embeddings and Clustering Parameters

3.4.4 Age & Gender Prediction

To complete the demographic profile for each identity cluster, age and gender were inferred from the representative facial images generated in the previous stage. Pretrained models from Hugging Face were used for this purpose, implemented through the transformers library (Wolf et al., 2020), with all code published in the author's repository (Wolf, 2025a; 2025b).

Gender was estimated with the SigLIP-based classifier developed by prithivMLmods (2024a). Each prediction was stored together with its confidence score and later merged back into the transaction-linked dataset. No filtering thresholds were applied at this stage, as the full distribution of confidence values was needed for later assessment.

Age estimates were produced using two independent pretrained models—prithivMLmods' CNN-based age detector (2024b) and the ViT-based classifier by nateraw (2024). The first model outputs the eight discrete age bins defined in this thesis, while the ViT model produces a raw prediction that is subsequently re-binned into the same categories. Outputs from both systems were saved to CSV and indexed via normalised filenames for alignment with the larger dataset.

As with any vision-based demographic estimation, results remain sensitive to lighting, pose, camera angle, and potential fairness limitations across demographic groups. These factors are addressed later in the Findings and Discussion chapters.

Attribute	Model Name	Model Type	Output Format	Processing Details
Gender	prithivMLmods/Realistic-Gender-Classification (prithivMLmods, 2024a)	SigLIP image classifier	Binary class + confidence	Inference via Hugging Face transformers; results annotated on images using OpenCV
Age	prithivMLmods/facial-age-detection (prithivMLmods, 2024b)	CNN with discrete age bins	8 age bins	Outputs labels per thesis-defined bins
Age	nateraw/vit-age-classifier (nateraw, 2024)	Vision Transformer (ViT) classifier	Raw age labels, re-binned	ViT predictions re-binned to 8 thesis bins

Table 4: Summary of Age and Gender Prediction Models

3.5 Evaluation Metrics

The final stage of the pipeline assessed how well the model-generated demographic attributes aligned with real survey responses and how useful these attributes were for downstream retail analysis. Survey data—captured on-site and stored in Excel—was linked to the accounting dataset through invoice numbers (Wolf, 2025e). After standardising column names and normalising free-text entries (including multiple survey submissions per invoice), the respondent with the highest demographic agreement score was selected. Their details were assigned to all linked invoice rows, creating survey_gender, survey_age_group, and a survey_completed flag for filtering. This alignment enabled the calculation of demographic

accuracy indicators without requiring separate ground-truth images. The key metrics used are summarised in the table below:

Metric	Description
Gender Match Rate	Percentage of invoice rows where predicted and survey gender align
Age Group Exact Match	Percentage of rows with exact matching predicted and survey age brackets
Age Group Approximate Match	Percentage of rows where predicted and survey age differ by ±1 bracket
Confusion Matrices	Visualization of classification performance and error patterns
Age Distance Plots	Visualization highlighting typical misclassification ranges

Table 5: Demographic Accuracy Metrics

Beyond accuracy validation, the merged dataset was explored for commercial insight (Wolf, 2025f). This included examining whether certain demographic groups showed stronger purchasing power, clearer product preferences, or distinct temporal patterns. The core analytical themes guiding these evaluations are listed in the table below:

Evaluation Category	Description
Confidence Analysis	Distribution analysis of prediction confidence scores for gender and age classifiers; high-confidence subsets analyzed for signal reliability
Revenue by Demographics	Analysis of average and total transaction values (AccoAmount) by predicted gender and age to infer purchasing power and identify high-value segments
Category and Product Preferences	Cross-tabulations of top-selling product categories and items by demographic groups; visualization using treemaps and Pareto charts
Temporal Patterns	Examination of hourly, daily, and weekly revenue patterns segmented by demographics to guide marketing and staffing
Survey Completion Behavior	Comparison of transaction amounts and demographic representation between survey-completed and non-completed subsets to identify potential labeling bias

Table 6: Business-Focused Evaluation Categories

Together, these evaluation steps offered both a technical read on the demographic model's performance and a practical view of how predicted customer segments behaved, providing a foundation for operational decisions and a clearer understanding of retail customer dynamics.

3.6 Limitations and Feasibility of Cloud Vision Analytics Infrastructure

Deploying cloud-based vision analytics at the POS level raises practical concerns around storage volume, hardware costs, and monthly cloud fees—issues that are especially relevant for SMEs. A single POS camera produces roughly 2 TB of footage per month, and storing this on Amazon S3 Standard typically costs around €40–45 per station (Amazon Web Services [AWS], 2023a; Security Camera Shop, 2023). Real-time processing using EC2 instances, such as t3.medium, adds an additional €30–60 per month (AWS, 2023b).

Initial hardware costs also remain significant. High-definition cameras, NVRs, PoE switches, and related networking gear—such as the Ubiquiti UniFi G4 line—bring the upfront investment to roughly €450–600 per station, with an additional ±€600 for site-level networking (Ubiquiti Inc., 2023a, 2023b, 2023c). Although AWS does not charge for inbound video transfer, the required upstream bandwidth (≈67 GB/day per camera) may introduce ISP-related costs depending on the store's connection.

A consolidated overview of these requirements is presented in Table 7, illustrating the combined capital and operational demands of a cloud-based setup:

Category	Item	Estimated Cost	Notes
Cloud Storage	Amazon S3 Standard (2 TB/month)	€40–45 per POS station	Scalable, pay-as-you-go
Cloud Compute	Amazon EC2 t3.medium instance	€30–60 per month	For real-time analysis; shared instance may reduce cost per station
Hardware	Ubiquiti UniFi G4 Camera & Accessories	€450–600 initial investment	Includes local NVR and PoE devices
Site Network Infrastructure	Routers, PoE switches, Wi-Fi access points	~€600 initial investment	Required for camera connectivity
Internet Bandwidth	Upstream capacity (~67 GB/day)	Variable	May incur ISP fees depending on store location

Table 7: Estimated Cloud Vision Analytics Infrastructure Costs and Requirements

From a feasibility perspective, SME affordability hinges on scaling strategies. Shared EC2 instances, for example, allow multiple POS stations to run on a single compute node—reducing per-station monthly costs and improving overall efficiency (Nutanix, 2023). Combining these factors provides a clearer picture of the financial and technical trade-offs involved in deploying cloud-based vision analytics in smaller retail environments.

Chapter 4 — Research Findings

4.1 Prediction Accuracy Results

Preprocessing the entire data pipeline — video surveillance data, POS transaction records, and image-based facial inference — yielded a dataset for downstream predictive modelling

and analytical functions. This led to fully traceable records that include rows corresponding to product-level transactions and an identified customer face, enriched with predicted demographic attributes.

After processing of CCTV and individual detection at the POS with YOLO, face cropping and enhancement by YuNet [optional super-resolution CNN (FSRCNN) or face restoration algorithm (GFPGAN)], facial images undergo gender and age prediction using transformer-based models developed from the Hugging Face platform. These predictive data were combined with matching POS records and the available survey data to create a unified structured dataset, the full schema of which is provided in Appendix C (Table C1: Final Dataset Schema After Preprocessing).

The AccoID and timestamp of each POS transaction link each image back to a POS transaction. The predicted gender and age were derived from facial analysis using Hugging Face transformer-based models, and transactional metadata was obtained from the store accounting system via SQL. Preliminary survey linkage fields are also included, though many records remain unmatched to participant survey responses (survey_completed flagged as FALSE). This structured dataset serves as the basis for all subsequent analysis and visualisations in the later sections.

The columns survey_time, survey_age_group, and survey_gender included survey_data, indicating that 61.1% of surveys were incomplete. The invoice number (InvNo) had about 2.5% missing data. A full visualisation of missing-data patterns is provided in Appendix C (Figure C2).

Of the entire dataset, 1,787 records included valid predictions for gender, age, and revenue metrics. Some 695 (39%) rows had corresponding survey data with completed demographic information. A summary of dataset validity counts—including valid demographic predictions, revenue-linked rows, and survey-completed entries—is provided in Appendix C (Figure C3:Data Validity Summary).

A scanned page of the manual survey sheets used for self-reported age and gender records at the POS was also provided in Appendix E, followed by a digitisation and alignment with invoice numbers and timestamps. The majority of missing demographic fields (~60%) derive from the incomplete participation captured in these manual records.

4.2 Prediction Accuracy Results

Model performance was evaluated by comparing the demographic labels verified by a survey with the predicted attributes; thus, a direct comparison of predicted and self-reported attributes was possible. Two age prediction models were applied, and the same gender classification model, prithivMLmods/Realistic-Gender-Classification, provided consistent gender predictions across models.

The prithivMLmods/facial-age-detection model performed best on most crucial metrics for reliable demographic segmentation of retail analytics. Accuracy was assessed on three dimensions:

- Exact age-group accuracy: percentage of predictions which precisely match the survey age category.
- Partial age matches: predictions falling within one adjacent bracket of the true age group, reflecting near-correct estimations.
- Overall accuracy: correct gender classification combined with either an exact or partial age-group match.

Taken together, these metrics account for both strict and relaxed correctness and collectively provide a holistic picture of model performance in real retail conditions. The comparison between the two models assessed in a structured manner is provided in the table below:

Metric	prithivMLmods Facial Age Detection	nateraw ViT Age Classifier
Gender Accuracy (%)	88.49%	88.49%
Age Accuracy – Exact (%)	18.99%	16.83%
Age Partial Match (%)	32.66%	44.46%
Overall Accuracy (%)	18.27%	14.24%

Table 8: Comparison of Age and Gender Prediction Performance Between prithivMLmods and nateraw Models

While the nateraw/vit-age-classifier model achieved a higher partial match rate, partial matches have relatively little analytical utility in the retail scenario because they obfuscate meaningful divisions that define which life stage is closest in proximity (e.g., if 31–40 is wrongly labelled as 21–30). In stark contrast to the above-mentioned exact-age accuracy gains, as well as the much higher overall accuracy of the prithivMLmods/facial-age-detection model, the latter may be better equipped to generate reliable demographic results correlating

with purchase intentions. As a result, all visualisations in Chapter 4 and the subsequent analytical results are generated solely from the prithivMLmods/facial-age-detection model selected.

Gender confidence scores indicate that most predictions have high confidence and cluster closely around 0.99. Even though a 0.5 confidence threshold was used to keep predictions, over 75% of all predictions exceeded 0.95 confidence.

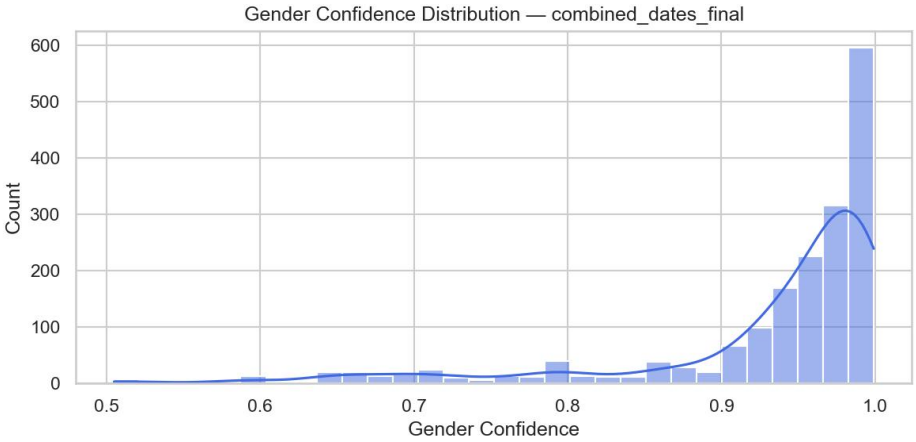

Figure 1: Gender Confidence Distribution

The boxplot analysis by predicted gender confirms consistently high confidence levels for both male and female classifications. Median confidence scores are 0.967 for males and 0.976 for females, with narrow interquartile ranges indicating low variability. Female predictions show a slightly higher central tendency, but both groups exhibit tightly clustered values within the 0.90–1.00 range and only a small number of low-confidence outliers, demonstrating stable and reliable model performance.

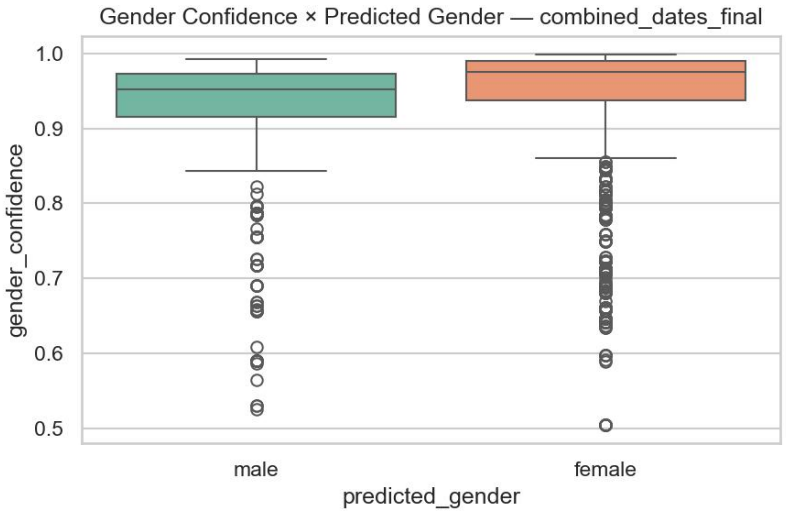

Figure 2: Gender Confidence by Predicted Gender

If they are confirmed against survey labels using a confusion matrix, they exhibit good correlation: 90% accuracy for predicted females and 83% for predicted males. The misclassification overall is still below 16%.

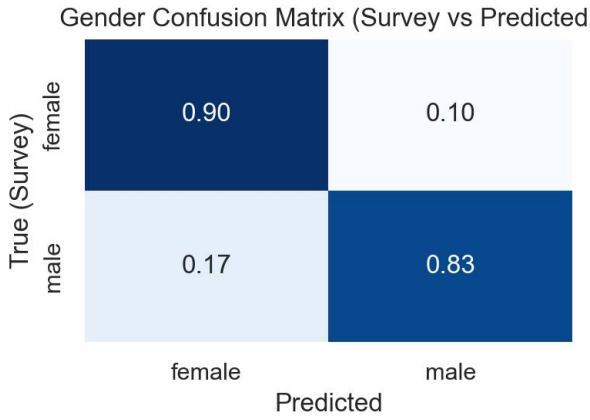

Figure 3: Gender Confusion Matrix

The histogram of age confidence scores indicates that most predictions fall within the range 0.40-0.70, with the highest bin centred at approximately 0.55, representing over 130 observations. The distribution shows moderate right skew, with fewer high-confidence predictions above 0.80. However, the central tendency suggests moderate confidence levels, indicating a lower degree of certainty in age predictions than in male and female predictions.

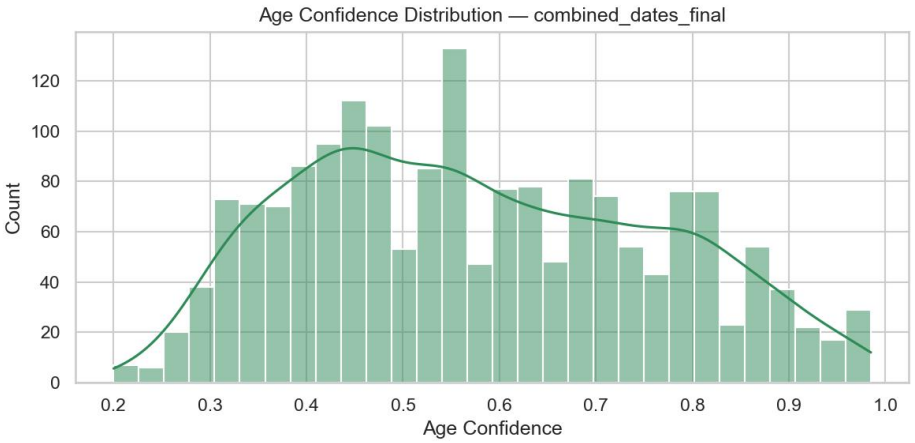

Figure 4: Age Confidence Distribution

Age confidence based on predicted gender: The age-confidence breakdown by predicted gender indicates that the interquartile range (IQR) for male predictions was smaller and the median confidence was higher (~0.58) than for female predictions (~0.54). Both groups showed considerable variation, with predicted values ranging from 0.2 to 0.98. This overlap demonstrates that the model's predictability was not rigidly stratified by gender.

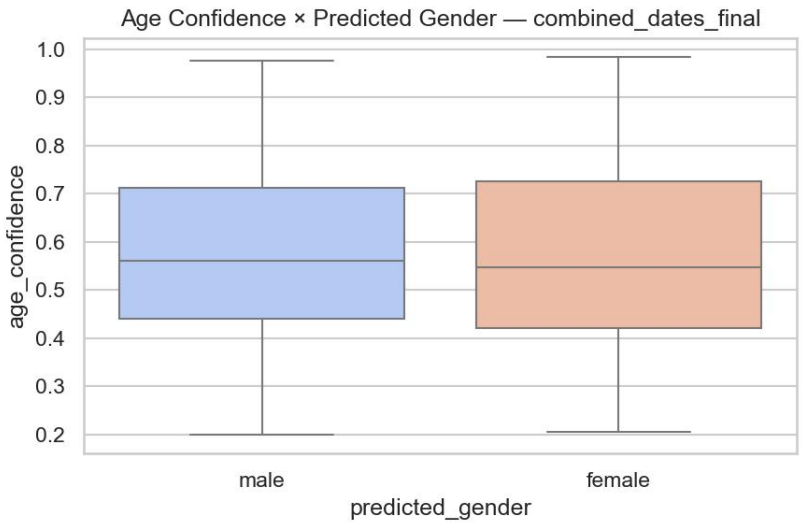

Figure 5: Age Confidence by Predicted Gender

Disaggregated by predicted age bracket, the median confidence for 11–20 was the highest (~0.72), while those for 31–40 and 41–55 were lower (~0.42–0.44). The range in predicted values for the 80+ group was wide, indicating highly uncertain predictions for this age bracket. Differences in these box plots indicate that prediction quality does not follow a standard pattern across age intervals.

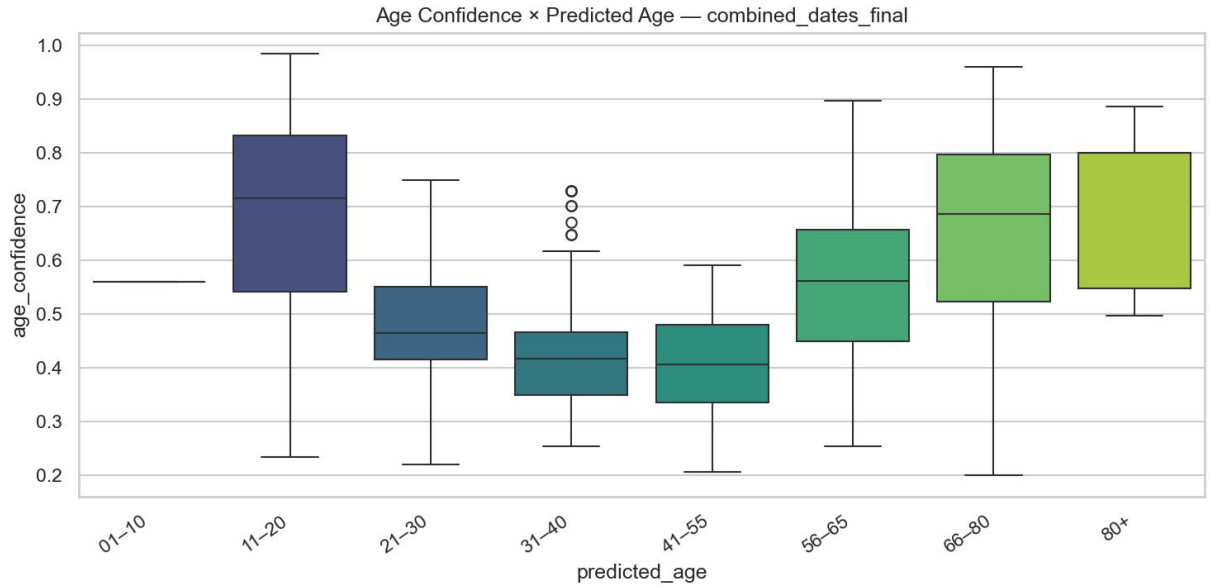

Figure 6: Age Confidence by Predicted Age

The confusion matrix shows the degree to which predicted and survey-reported age brackets agree. The greatest alignment is observed for 56–65 (55 accurate predictions) and 66–80 (59 accurate predictions). The 41–55 bracket has a moderate match rate of 33 correct

predictions. Notably, 68 participants in the 31–40 bracket were misclassified as 11–20. These patterns mirror significant cross-bracket confusion across some age ranges.

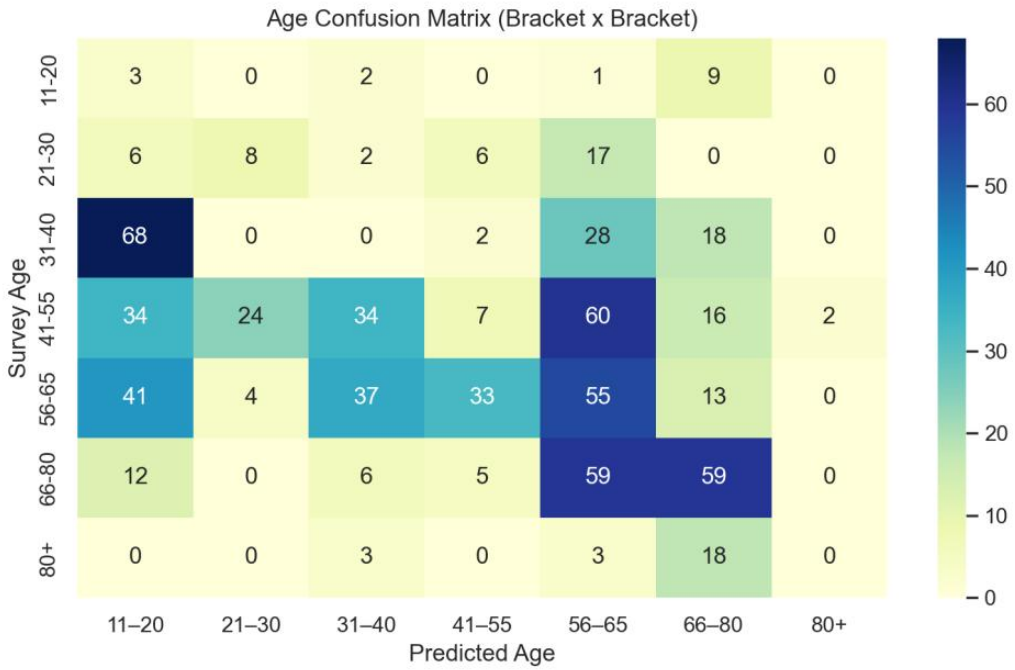

Figure 7: Age Confusion Matrix

The average magnitude of age-bracket errors (see Appendix C, Figure C4) shows that most misclassifications fall within 1–2 adjacent age brackets, with larger deviations occasionally observed for the 11–20 and 66–80 age groups. A breakdown of prediction error types shows that only 17% of age predictions are exact matches, around 45% fall within ±1, and the remainder exceed ±1 (see Appendix C, Figure C5). A joint-density visualisation of gender and age confidence values confirms that gender confidence remains tightly concentrated near 0.95, while age confidence is more widely distributed (Appendix C, Figure C6).

4.3 Demographic Distribution Results

Transaction amounts are distributed similarly for both predicted male and female customers (both with median transaction values close to 200 ZAR). The interquartile ranges are comparable, although the males' spread is slightly wider than the females'. Both groups feature a number of low-value, outlying observations, and the highest-value transaction is made by females at approximately 1350 ZAR. The overall central tendency across genders is stable and generally similar.

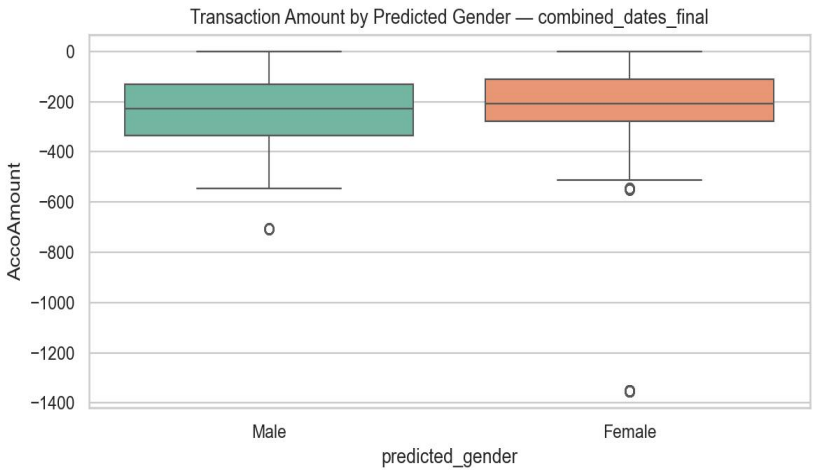

Figure 8: Transaction Amount by Gender

For most predicted age groups, transaction amounts also look very nearly similar. Median values range from 150 to 300 ZAR, and the interquartile ranges indicate a moderate spread. There are also isolated low-value anomalies across several age groups (specifically 31–40, 56–65, 66–80)—some ranging from 600 to 850 ZAR. The 80+ group is the only exception with a highly pronounced trend towards the tail end of the distribution, driven by a single extreme outlier near 1350 ZAR, which creates a very long lower whisker. Aside from this oddity, no age group shows significantly higher or lower spending than the others.

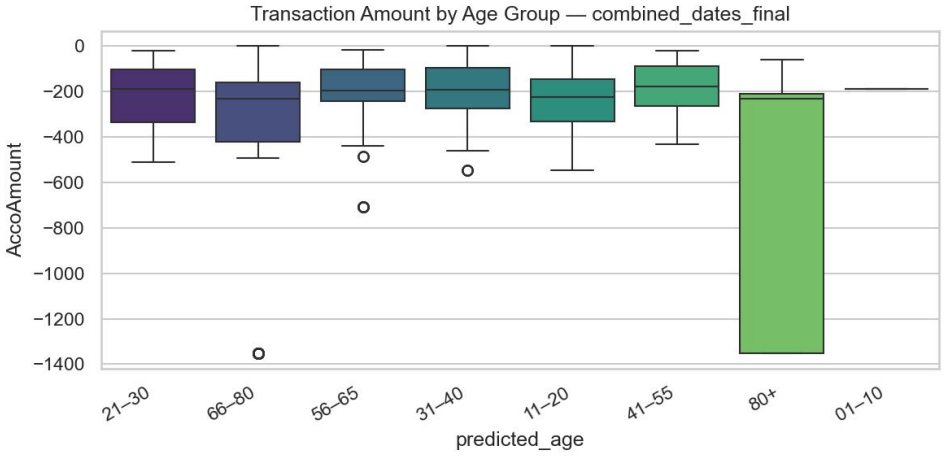

Figure 9: Transaction Amount by Age Group

Females generated much greater revenue, suggesting higher average transaction spending than males. The highest total revenue came from the 11–20 age group, followed by the 56–65 and 66–80 brackets. There was a notable drop in revenue among the youngest (01–10) and oldest (80+) groups, suggesting low participation in these demographic segments.

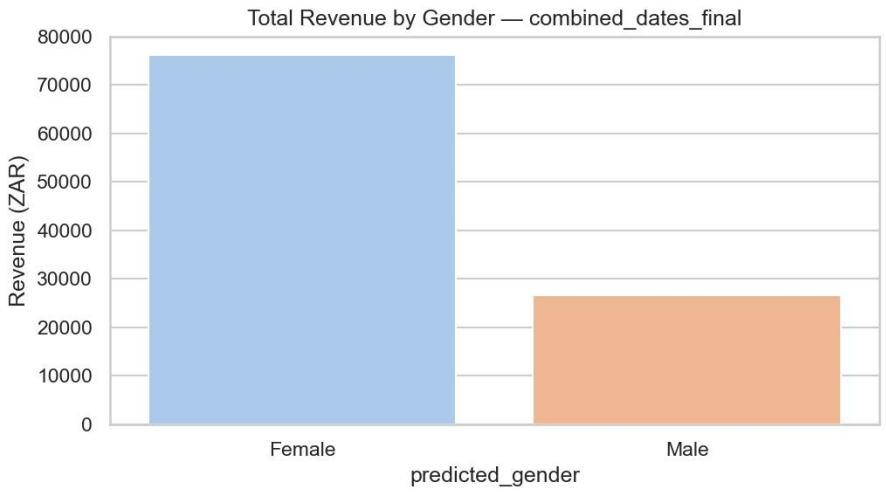

Figure 10: Revenue by Gender

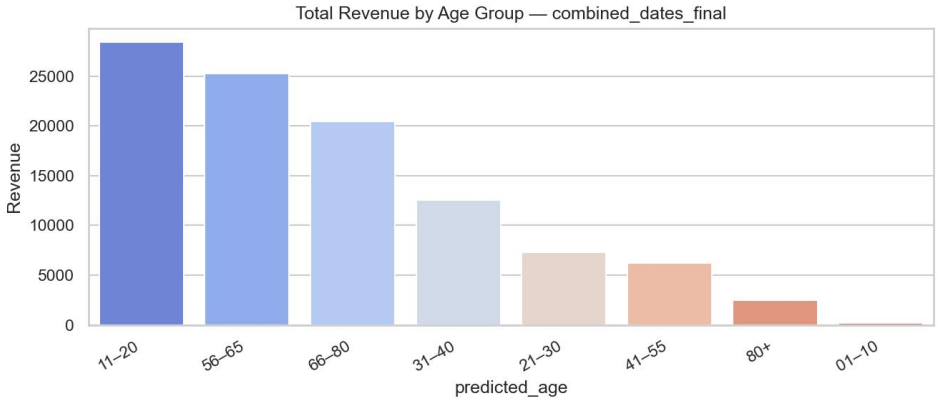

Figure 11: Revenue by Age

In the revenue heatmap, income composition changes considerably across gender–age groups. Female revenue contributions: 20,411 ZAR from females aged 11–20, 17,878 ZAR from females aged 66–80, and 16,341 ZAR from females aged 56–65. The majority of male revenue is lower across nearly all age groups, with the highest male contribution in the 11–20 range (8,003 ZAR). Transactions involving the youngest (01–10) and oldest (80+) groups are also rare, as these two groups play a limited role in the dataset.

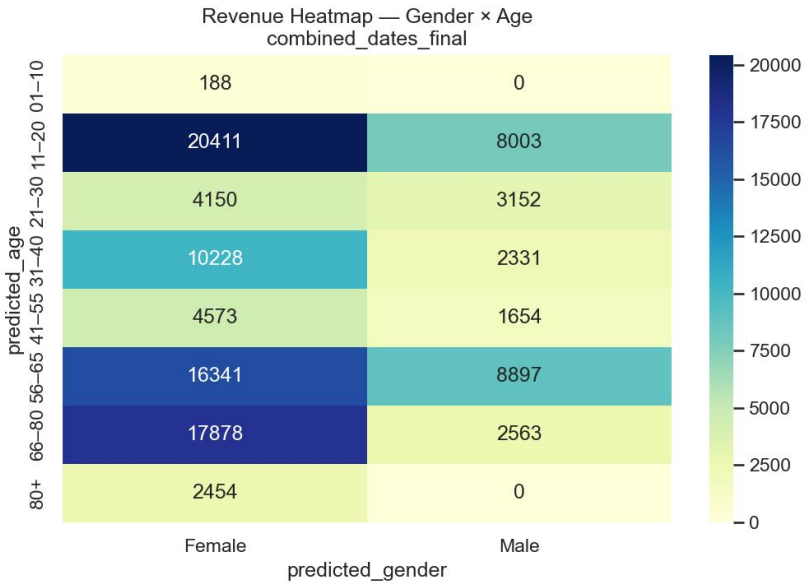

Figure 12: Heatmap of Revenue by Gender and Age

4.4 Revenue & Transaction-Based Analytics Output

Initial exploratory analysis of product categories showed a highly concentrated revenue structure, with a small set of categories (notably Beverages Coffee, Breakfast, and Light Lunches) contributing a disproportionate share of sales. Detailed category-level revenue and frequency plots are provided in Appendix D (Figures D1–D3) and are not repeated here to keep the focus on demographic patterns.

The remainder of this section, therefore concentrates on gender- and age-segmented revenue and transaction patterns, which directly address the research question on whether demographic insights derived from computer vision can provide actionable value for SMEs

As shown in the grouped bar chart, female customers were more profitable than male customers across most product categories. Females' highest contributions are visible in Breakfast, Beverages, Coffee, Light Lunches, and Cakes, where each is clearly greater than male totals. In general, male revenue is lower across categories, and contributions are only modest in most segments. In this chart, categorical revenue is determined more by female purchasers.

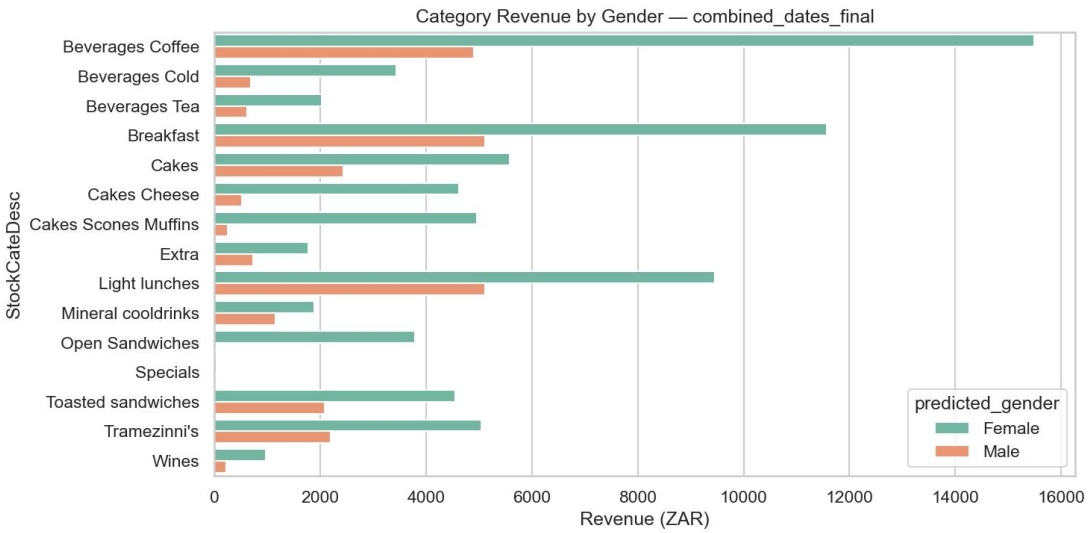

Figure 13: Category Revenue by Gender

Similarly, the grouped bar chart of purchase counts shows that, across almost every product category, females make more purchases than males. The most significant difference is observed in Breakfast, where female purchase counts are considerably higher than in any other category. Females lead in Beverages, Coffee, Light Lunches, and Cakes, although men can make a smaller but stable contribution in each category. These trends suggest that female customers account for the highest transaction count.

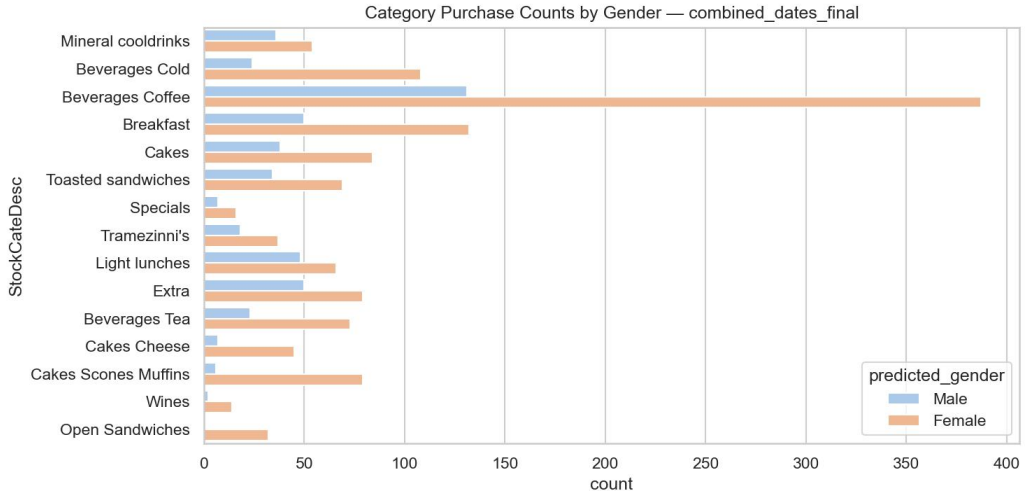

Figure 14: Category Purchases by Gender

The heatmap clearly shows distinct trends in concentration for the ten most purchased items. Cappuccino is the most popular beverage across all ages, with the highest counts in the 66–80 grouping (87), vigorous product volumes in the 56–65 grouping (50), and a peak activity in the 11–20 grouping (42). The one listed as “none” also shows peaks among older age groups, most notably those aged 56 to 65 (56) and 66 to 80 (45). Across all ages, other products show relatively low counts, with the youngest (01–10) and 21–30 groups showing

little activity. Things like Chicken Feta Basil Pesto and Tea Five Roses are still few and far between in every category. Demand is primarily for beverages for this age group.

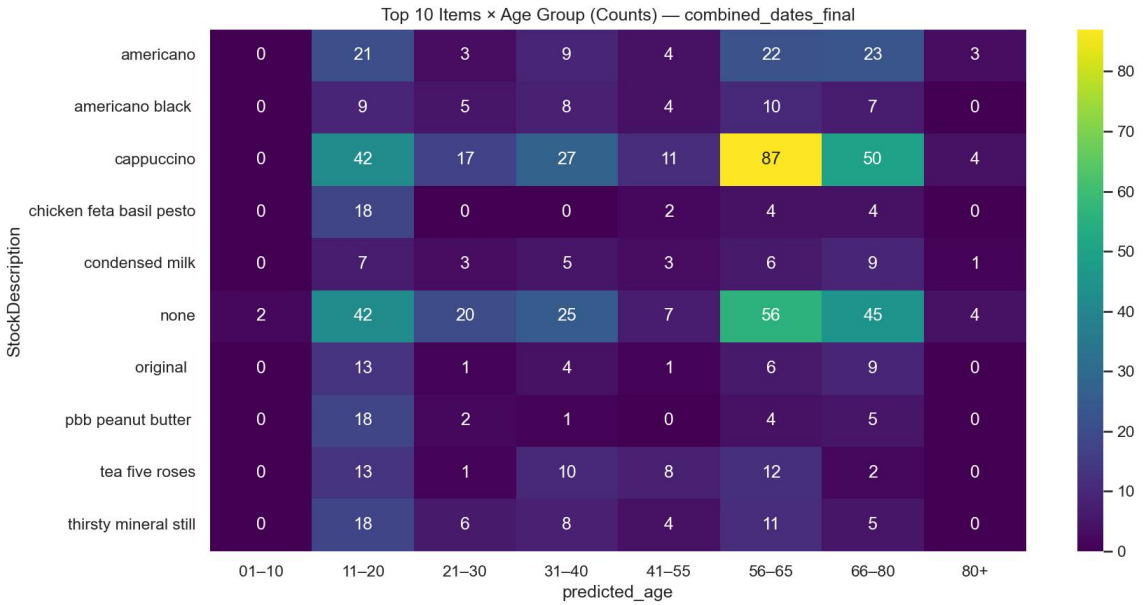

Figure 15: Top 10 Items by Age Group

The distribution of line-item count indicates significant variability in purchase behaviour over time. The activity is highest among 11- to 20-year-olds (~480 items), followed by 56- to 65-year-olds (~450) and 66- to 80-year-olds (~340). Middle-aged groups have a moderate level of participation: 31–40 at ~230; 21–30 at ~135; 41–55 at ~115, respectively. 80+ group activity is minimal (~30), and 01–10 group activity has been virtually negligible. On balance, purchase frequency is more concentrated among younger teens and older adults.

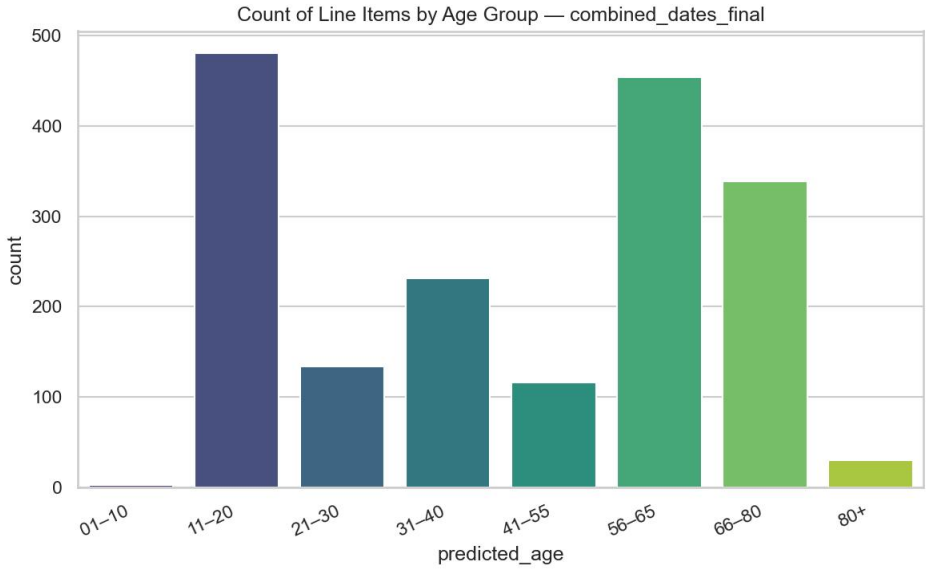

Figure 16: Count Line Items by Age Group

The Revenue distribution among age groups is highly centralised in certain cohorts. Of the total revenue, 11–20 groups yield just over 28,500 ZAR, while 56–65 (~25,200 ZAR) and 66–80 (~20,000 ZAR) do better. Mid-aged groups contribute significantly less: 31–40 contribute ~12,500 ZAR, 21–30 contribute ~7,500 ZAR, and 41–55 contribute approximately ~6,200 ZAR. The 80+ only accounts for low revenue (~2,500 ZAR), and the 01–10 group contributes almost no revenue. In general, revenues are skewed towards late-teen and older-adult audience segments.

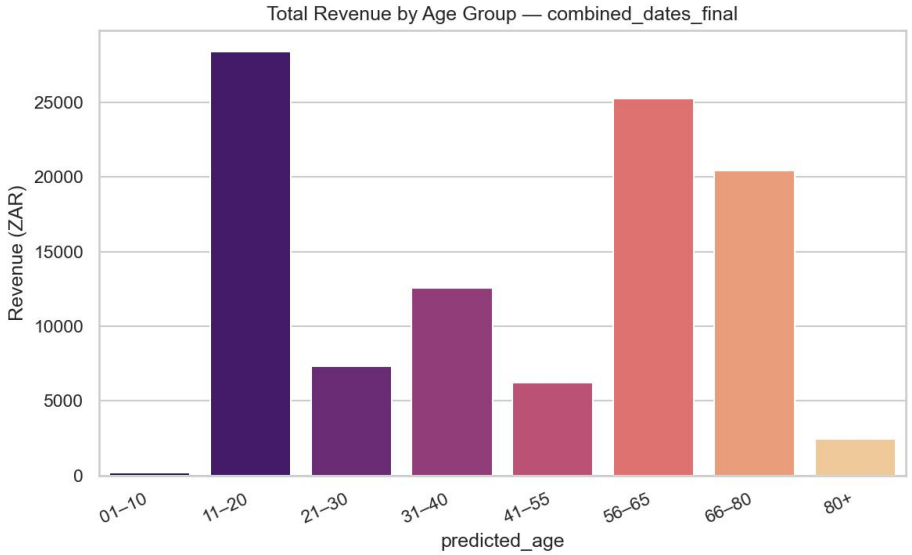

Figure 17: Total Revenue by Age Group

Additional exploratory plots — demographic distributions, extended item-mix heatmaps, and revenue-versus-confidence views — are provided in Appendix D. These supplementary visuals offer finer-grained insights into item- and category-level structures, detailed revenue contributions, and more granular segmentation of purchasing behaviours by predicted age and gender. Also in the appendix, alternative category-share summaries and confidence-based analyses demonstrate how the risk of incorrect predictions varies across products. Collectively, these materials illuminate the breadth of exploratory results without distracting from the main chapter's focus.

4.5 Temporal Patterns in Revenue and Transactions

Revenue and transaction activity are concentrated almost entirely in the morning and early afternoon, with clear peaks between 09:00 and 13:00. Only two trading days were captured in the dataset, so all weekday patterns reflect Wednesday and Friday activity. Overall revenue across the two days remains stable with only a slight dip on the second day. All supporting visuals for these temporal patterns are provided in Appendix D4–D8.
Male transaction counts are lower than female transaction counts in every active hour. Female activity increases from about 30 transactions at 08:00 to a peak of around 280 at

10:00, with later peaks near 220 at 12:00 and 13:00. Men start trading at about 30 at 08:00, peak in about 85 at 10:00 and then oscillate between 30 and 110 until midday. After 14:00, they both have relatively low activity.

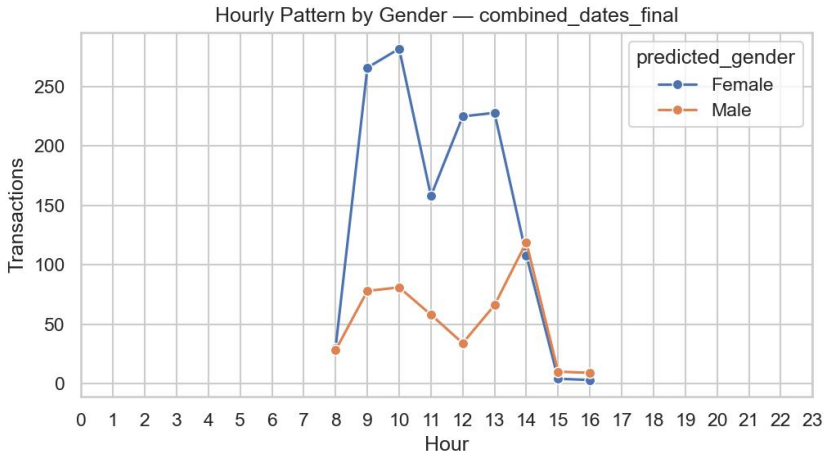

Figure 18: Hourly Transaction Pattern by Gender

All age groups recorded transactions exclusively through Wednesday and Friday, as per the two-day recording window. The highest activity levels are observed in the 11–20 and 56–65 groups each day, with a peak of about 270–280 transactions. Moderate volumes are observed for the middle-aged groups, including individuals aged 21–30, 31–40, 41–55, and 66–80, with 50-170 transactions per active day. On both days, the transaction counts from the groups of 01–10 and 80+ remain consistently low.

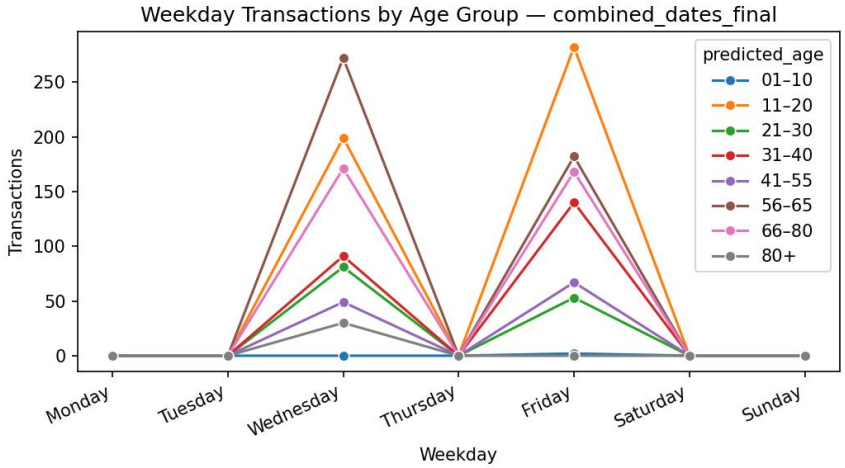

Figure 19: Weekday Transaction by Age Group

A breakdown of hourly transactions shows different concentration profiles between age groups. The 11–20 group shows the most activity, with a peak around 115 transactions between 10:00 and 11:00, and volumes remain high from 12:00 to 14:00, with counts ranging from 70 to 88. Mid-morning clustering is also observed in the 31–40, 41–55, 56–65, and 66–

80 groups, with peaks at 50–115 transactions. Moderate-to-high engagement behaviour is observed for the 21–30 group at 15–30 transactions per hour. The 01–10 and 80+ groups in particular have very limited participation, with only sporadic low-value observations.

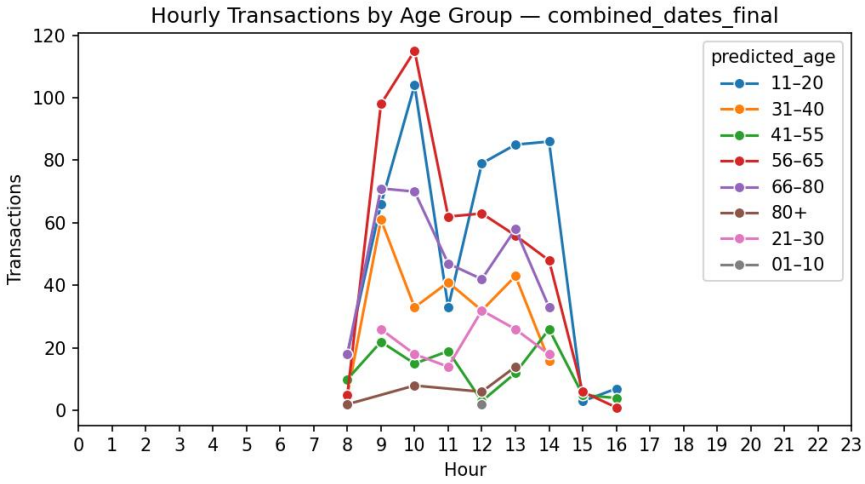

Figure 20: Hourly Transactions by Age Group

Trends in revenue by age group mirror those in the transaction tables. The highest revenue levels are observed in the 11–20 and 56–65 groups, with peak values of approximately 6,900 ZAR and 6,100 ZAR, respectively. The middle class contains 66–80 (~3,800 ZAR), 31–40 (~2,700 ZAR), 21–30 (~1,600 ZAR), and 41–55 (~1,000–1,200 ZAR), with a revenue at a rate comparable to that of the transactions. Poor participants, like 01–10 and 80+, on the lower end will always generate under 1,000 ZAR per hour.

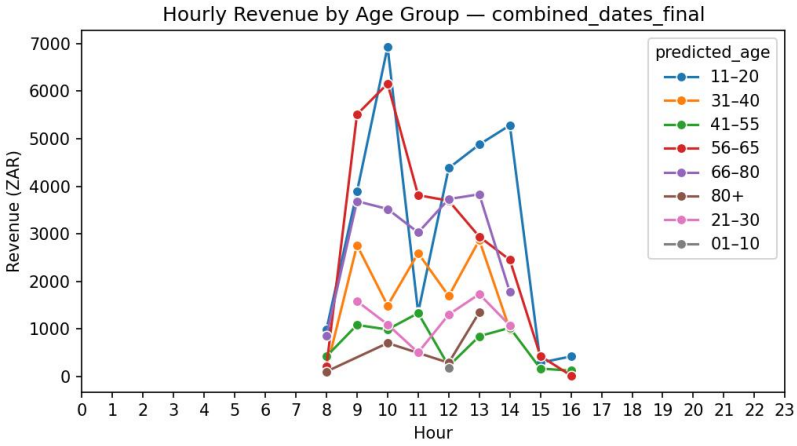

Figure 21: Hourly Revenue by Age Group

Chapter 5 — Discussion

5.1 Alignment with Literature

The results in Chapters 3 and 4 fit well with existing literature on computer vision-based demographic inference, embedding-driven identity representation, YOLO detection, and SME

capacity constraints. Chapter 2 addresses the broader context, while this section discusses how the proposed pipeline resembles patterns observed in previous research.

As anticipated, gender prediction outperformed age estimation in terms of accuracy, in line with evidence suggesting a good balance of gender under a variety of conditions; age prediction still depends on lighting, pose, and natural differences in faces. The younger skewness of predictions mirrors common misclassification problems found when models trained on curated data are generalised across uncontrolled CCTV data.

YOLOv11 use is in line with the literature on the strengths for real-time detection, especially under cluttered conditions. Practical adaptations — such as bounding-box filtering to avoid occlusions and erratic shopper movement — mirror issues that previous research has repeatedly highlighted. FSRCNN enhancement and OpenCV-based preprocessing were also applied as proposed techniques for stabilising CCTV-quality images.

Identity grouping using embedding also mirrors previous studies to a very high degree. Application of FaceNet-style embeddings and density-based clustering yielded the expected behaviour: stable clusters when lighting and pose were stable, and fragmented when they were not. Iterative merging, coupled with outlier handling, captures the real-life complexities that are missing from benchmark datasets.

The paper highlights an even more unique aspect of that research: it can integrate demographic data derived from CCTV with POS transaction records in an SME context—an integration generally lacking in current reviews, which tend to treat these aspects of the information in isolation or cover only large vendors. Proportioning an end-to-end pipeline from detection to demographic attribution to transactional alignment demonstrates that SMEs can implement ways of working typically developed within an entity of a larger scale and a larger dataset.

Ultimately, the concrete limitations faced—data sparsity, consumer-grade cameras, and limited hardware—resonate with established gaps in SME digital capability. Furthermore, the use of lightweight models and open-source tooling, complemented by POPIA/GDPR compliance management, adds a practical aspect that is often overlooked in theoretical debates. The results add to the literature about performance, bias, and robustness and better specify the behaviour of these approaches with SME businesses.

5.2 Feasibility for SMEs

The results show that SMEs can, in fact, extract usable demographic and behavioural insights from existing CCTV systems without relying on enterprise-grade infrastructure. The work in Chapters 3 and 4 confirms that common SME limitations—older cameras, uneven lighting, and modest hardware—do not prevent the pipeline from functioning, but they do shape how the system must be deployed and maintained.

The technical workload remains a central consideration. Although the workflow is built entirely on open-source tools, tasks such as embedding generation, clustering, and image filtering remain computationally intensive. Even processing two days of footage from a single camera illustrates the limits of typical SME hardware and aligns with the broader literature, which notes constrained processing power and storage. While periodic or batch processing is manageable, continuous multi-camera operation generally requires either cloud support or scaling strategies that can grow as demand increases.

Data quality challenges from the physical environment are equally important. The overhead camera provided sufficient visual detail for detection and cropping, but inconsistent angles, lighting shifts, motion blur, and occlusions reduced age-prediction accuracy and occasionally led to cluster fragmentation. These issues are expected in SME CCTV footage and highlight the value of strong preprocessing, super-resolution, and strict filtering rather than undermining feasibility.

Linking visual data with POS transactions also determines how realistic adoption is. SMEs often run separate systems—CCTV, POS, accounting—that do not integrate cleanly. The timestamp alignment, consistent file structures, and manual corrections required in this study mirror those typical fragmentation issues. Yet the successful end-to-end linkage shows that SMEs can approximate e-commerce-style insights without major system changes when data is carefully structured.

Affordability remains favourable. Most SMEs already operate CCTV setups, and devices such as UniFi G4 cameras, NVRs, and mid-range GPU workstations fall within typical upgrade budgets. Cloud costs, especially under pay-as-you-go models or shared processing across multiple POS stations, make ongoing deployment practical and challenge older assumptions that such analytics were financially out of reach for small retailers.

Finally, feasibility depends on the lawful handling of data. By deleting raw footage, storing only anonymised embeddings, and linking visual outputs strictly to transaction identifiers, the pipeline aligns with key principles of POPIA and GDPR. Specifically, this approach ensures compliance with data minimisation as outlined in Article 5 of the GDPR. Such privacy-by-design choices bolster trust and mitigate risks associated with biometric data usage, further adhering to the limited use doctrine (POPIA, 2013).

5.3 Comparison to E-Commerce Analytics Benchmarks

The goal of the study was to analyze whether SMEs can derive intelligence from CCTV-based computer vision that is consistent with the intelligence e-commerce platforms commonly provide. Although physical retail never comes close to digital clickstream metrics, Chapters 3 and 4 demonstrate that aspects of online behavioral analytics are transferable to a physical store, albeit through different means and with different constraints. In-store analytics mainly rely on indirect inference from camera footage. Yet, the pipeline still yields usable demographic and behavioral signals from direct observation.

E-commerce is a winner because every interaction with the customer is explicitly logged: searches, page views, recommendations, and conversion paths are all logged. Age and gender predictors, while they do not achieve complete accuracy, do permit segmentation of revenue and transactions. High-value customer groups that reflect the patterns of individual products actually purchased can be identified through a demographic lens. The result is on par with any digital dashboard, but it is based on visual inference rather than online tracking.

For instance, a retailer could discover that female customers predominantly shop in the morning, allowing them to allocate more staff during these peak times to enhance service efficiency. Furthermore, understanding age-based buying habits enables targeted promotions that cater to specific demographics, increasing conversion rates. This active application of insights shifts analytics from mere observation to strategic action, illustrating how physical retail can respond dynamically to consumer behaviors

Through the integration of CCTV outputs with POS data, SMEs have the following analytic functionality, which is typically only available via online platforms: category-level performance by demographic segment, item revenue distributions and time-of-day behaviour patterns by age and gender groups. While online systems track customers across stages of a session, this pipeline provides a one-time view of a user session by removing raw imagery (in

compliance with POPIA and GDPR) and by not using persistent identifiers. That's the starkest divide between physical and digital analytics.

The overlap is further strengthened by temporal behaviour analytics. The hourly patterns of activity, where your demographic peaks are, the revenue cycle and such serve like e-commerce demand forecasting and campaign timing software, but are built on these smaller timeframes of use and the rarity of an event in a store. The main constraints are related to attribution: online systems are connecting the dots to campaigns and referral sources, while CCTV cannot. Even so, the demographic outputs are empowering offline strategies that many SMEs cannot match through direct footfall counts or anecdotal observation.

The correspondence is strongest at the product and category levels. Integrated CCTV–POS analysis generates demographic-driven item heatmaps, revenue hierarchies, and purchasing profiles, providing SMEs with insights that previously relied on enterprise systems and mirroring an e-commerce analytics framework in its structure. These observations illustrate that computer vision can bring some of the strategic insight of online retail to physical retailers at low cost and with minimal organisational change.

Above all, it can be said that CCTV-based analytics can never replace e-commerce's longer-term tracking and the ability to link purchases to different marketing touchpoints. However, they are a great way to mimic core analytical layers—segmentation, behavioural profiling, temporal trends and category insights—alongside the physical retail setup and privacy measures. For SMEs, it's a reasonable solution, an attempt to reduce the distance between the online sophistication and offline experience from a privacy viewpoint.

5.4 Data Quality & Model Performance Analysis

Input data and the behaviour of AI techniques for detection, cropping, embedding, and demographic inference directly shape the pipeline's insights quality. In Chapters 3 and 4, we demonstrated that the system is consistent for SMEs even under common resource constraints, and that the results indicate how environmental conditions and model behaviour can impact demographic accuracy and downstream analytics.

The dataset, consisting of two days of static POS camera footage, is representative of the short-term, episodic sample style common in SMEs, where maintaining and continuously recording an event may be impractical in the long term. This finite time span accounts for the temporal gaps identified in Section 4.5, but, again, while short samples can provide

meaningful behavioural signals when processed systematically, even small samples may still represent meaningful behavioural cues.

In an overhead, mixed lighting setup, data quality issues were anticipated. Shadows, obstructions, and partial angles impeded the picture and reduced clarity, as mentioned above, especially in predicting age. SigLIP-based gender classification was retained to a good degree, consistent with reported >95% accuracies in stable environments, and gender classification performed similarly in stable settings, but age estimation often reported misclassifications across distant age bins, a common event in uncontrolled CCTV images. These findings reflect well-established limitations of age estimation models that perform across a range of illumination, pose, and movement.

There were also factors influencing evaluation quality by survey alignment. Only 39% of transactions contained survey responses, and a rule-based choice of the “most likely” respondent introduced uncertainty. However, gender predictions closely matched survey data, and age comparisons showed the expected variability. This asymmetry is indicative of both model behaviour and the imperfection of human-reported “ground truth” for age.

The embedding and clustering performance were also very sensitive to face-crop. Facenet512 embeddings yielded coherent identity clusters, but noisy or partially cropped images led to noise points in HDBSCAN. Fragmentation decreased with centroid-based merging and outlier reassignment, but certain levels of uncertainty were persistent. Crucially, cluster exemplars were generally clear enough for valid demographic inference and steady association with POS transactions.

These technical issues didn't stop useful analytics. Combining clustered identities with transaction data enabled granular revenue segmentation, category–gender segmentation, and behavioural profiling. The pipeline's worth lies more in a direct, reliable linkage between face and transaction than in perfectly predicted demographics. Timestamp matching proved highly successful, indicating that SMEs can perform action-based segmentation without the continuous user identifiers used in e-commerce applications.

Age-prediction uncertainty does not detract from the utility of the analytic approach. The results act as aggregate, probabilistic signals and not as truths in isolation – both technically and ethically correct. Although single-model predictions require caution, cohort-level trends are directionally robust and can be used in common applications (e.g., classifying high-value demographics, preference categories, or peak trading periods) within the SME industry.

Taken together, the findings reflect the true conditions in SME environments – limited recording windows, mixed lighting, partial occlusions, and insufficient ground-truth labels. Model performance is in line with previous literature – good gender accuracy, moderate and varying age predictions and consistent embeddings under non-ideal conditions. Although data were imperfect, the pipeline provided intelligible demographic and behavioural insights, suggesting its practicality for SMEs with present CCTV infrastructure.

5.5 Insights for Retail Marketing Strategy

Chapter 4 demonstrated that even a very low-end CCTV system can extract demographic and behavioural information that can be directly used in retail marketing decision-making. Conventional sales totals for SMEs and manual observation are hard sell, but linking demographics to item-level transactions is a step toward the kind of segmentation we typically encounter in e-commerce.

Clear strategic themes emerge. High-volume categories (Beverages: Coffee, Breakfast, and Light Lunches) account for the majority of revenue and foot traffic, underscoring their positioning as habitual, low-friction products that serve as an anchor point to their daily store routine. Females consistently generate the most revenue across categories, with a significant share in high-frequency items, reflecting overall retail evidence that finds women to be the main contributors to everyday food-service spending. SMEs can leverage this through promotions, in-store advertising messages, and loyalty incentives specifically for female shoppers.

Targeting becomes even more refined through age-based insights. We see good activity from those around 11–20, which translates to a consistent, repeatable customer segment. In the 80+ group, it's fairly small, yet they all have higher basket values—characteristic of those who make premium or convenience-driven purchases. This underpins differentiation: high-turnover items and quick-service offerings for younger shoppers vs. bundling, convenience positioning, or premium categories for older shoppers.

Temporal patterns mean something in terms of operations. The pronounced morning and lunchtime peaks highlight opportunities for time-specific promotions and more efficient staffing and preparation schedules. SMEs may not have real-time systems or dynamic pricing, but, as you can see, the market size depends on demographics. Only because CCTV-derived demographics are linked to POS transactions does this sort of insight exist—

something SMEs typically miss without loyalty programs. Although the age estimates are imperfect, they are directionally valid, and gender estimates are very accurate. This is sufficient to support segmented marketing, assess category performance by demographics, and monitor purchasing behaviour over time.

As a whole, integrating CCTV analytics and POS data helps SMEs move from intuitive to targeted, evidence-based marketing. These results are consistent with the retail analytics literature on gender category trends, age-related basket differences, and time-based demand cycles, suggesting that, in a relatively short period, even data-driven patterns emerge. Longer-term use and ongoing refinement are needed for SMEs to catch up with data-driven optimisation, which is common among larger retailers and online providers.

5.6 Realistic Operational Considerations for SMEs

While the insights generated by the CCTV–POS pipeline can be useful, deploying them in real-world settings can be more challenging than in a controlled research setting. For instance, small South African retailers experience operational pressures on a day-to-day basis that affect system reliability.

One perennial challenge is the erratic nature of CCTV infrastructure. Many SMEs have poorly maintained cameras, frequent misalignment, and interruptions due to load shedding. These factors, in combination with changes in lighting, drifting and occlusion, impacted the model's confidence on a large scale throughout the entire study. Although techniques such as FSRCNN, GFPGAN (later disabled), and YuNet filtering were useful for stabilising inputs, the computational requirements for these adjustments frequently exceed SMEs' capacity to implement them in practice. This reflects broader findings indicating that hardware is fragile, and that a lack of technical expertise can put off using AI in ways that algorithms never did before algorithms become a limiting factor.

Staffing shortages complicate matters further. Employees currently handle several obligations, which limit the resources for system maintenance. The manual timestamp reconciliation in this investigation is impractical at scale. Demographic–transaction linkages are rapidly losing reliability, with no automated clock synchronisation, camera health monitoring, or direct POS integration.

In practice, the system needs to be run continuously and invisibly, with no manual intervention. These difficulties are exacerbated by infrastructure constraints. Load shedding

disrupts recordings and cloud sync, and poor internet connectivity hampers cloud inference. While off-device processing models are helpful, cloud processing remains susceptible to latency, variable availability, and the volatility of monthly costs. And local GPU inference would help mitigate these risks, but it would require hardware, cooling, and power that many SMEs cannot provide.

This makes managing cloud costs a major concern. Costs can bill unexpected spikes in activity during peak-hour inference times. High-resolution face enhancement slows processing and increases the volume of cloud calls, with lower margins. For SMEs, therefore, predictable cost ceilings, batching systems, or lightweight on-device inference are required, though even lighter models come at the cost of accuracy.

Even when the analytics show us useful patterns (morning peaks, gendered category preferences, high-value elderly baskets), operating on those patterns demands operational flexibility. Most SMEs are not equipped with the necessary capacity or process structure to change product placement, stock preparation or offer tailored promotions on the fly. The system is most useful if it enables simple-to-manage actions rather than requiring real-time tactical fine-tuning.

POPIA also determines whether something may be feasible. Its pipeline achieves this by deleting raw video and storing only anonymised embeddings, but sustainability depends on transparent communication, governance, and data-handling protocols. Many SME managers have no legal help/support, leaving them on higher alert for unintended non-compliance or customer concerns - the problem arises whenever a practice is not used well enough.

These realities, combined, point to SME deployment being possible only in the automation-first scenario. This needs proper camera placement, automatic time synchronisation, offline batched inference, and a fully unattended flow linking CCTV data and POS transactions. Any implementation based on manual upkeep or human real-time oversight will likely add more friction than value. The research thus demystifies both the opportunities and the limitations in this regard: SMEs can attain a certain level of demographic intelligence regarding e-commerce, but only with stable infrastructure and thoughtful automation, given their resource limitations.

Chapter 6 — Conclusion

6.1 Summary of Research

This work investigated whether SMEs with small budgets, basic CCTV, and no customer registration systems can derive useful demographic and behavioural information comparable to that available in e-commerce contexts. The study proposed an inexpensive pipeline that includes person detection, face cropping, embedding-based identity grouping, and open-source demographic inference models, and tested it on 2 days of CCTV-enabled POS footage from a South African retailer. The pipeline reliably produced accurate gender predictions and directionally useful age estimates, even under nonideal camera angles and varied lighting conditions. If these inferred properties were associated with transactions, we were able to clearly segment the revenue, purchase and temporal activity patterns. It thus enables SMEs to reconstruct a demographic layer from point-of-sale information using only their currently running CCTV, assuming adequate data quality and automated processing. All key research aims were met: demographic inference from CCTV is practical from a technical standpoint; the features support useful analytics; the system is SME-friendly; and there are indications that the findings will approximate several key components of e-commerce customer intelligence.

6.2 Final Evaluation of SME Feasibility

These findings suggest that SMEs can implement demographic analytics using CCTV-based systems, but this is feasible only if the platform operates stably. Those minimum requirements include fixed, well-positioned cameras, consistent lighting, accurate timestamp alignment between devices, and automated processes without staff intervention. In such circumstances, SMEs may enrich POS data with demographic data and support strategic decisions about product mix, staffing schedules, and marketing focus. However, feasibility is conditional rather than guaranteed. Power instability, uneven internet quality, camera drift, and limited technical expertise, all common in the South African SME setting, directly influence data quality and system reliability. Thus, the model is most effective when ingestion, synchronization, and processing can be automated without manual correction. With these safeguards in place, the system provides SMEs with a practical way to gain structured behavioral and demographic insights using the infrastructure they already possess.

To facilitate immediate action, consider rolling out the system in stages. A tangible milestone for the first month might be achieving an 80% demographic linkage using CCTV data from one camera. This goal not only sets a clear benchmark for early success but also helps in

building momentum and confidence in the phased deployment plan. This structured approach allows SMEs to evaluate system efficacy and refine operational tactics before full-scale implementation.

6.3 Contributions of This Study

First, it presents a methodological tool as a single, unified, step-by-step pipeline that connects CCTV-based person detection, embedding-driven identity grouping, and open-source demographic inference with POS transaction data. This combination, suitable for resource-limited environments, is not demonstrated in the literature. Second, it is an empirical test that illustrates how even short observation windows via CCTV can yield stable demographic and behavioural indicators that correspond to what we recognise as retail behaviour. That everyday security cameras—which are already found in most SMEs—carry latent analytical value well beyond their role as a security resource. Third, the study provides an actionable SME contribution by setting a replicable approach that allows small retailers to estimate certain dimensions of e-commerce customer intelligence without loyalty programs, specialist hardware, or technical staff. It elucidates what insights are realisable and where operational constraints prevent analytic depth. Last but not least, the work also advances ethical governance by offering a privacy-preserving configuration that uses anonymised facial embeddings, deletes raw facial images, and aligns with POPIA and GDPR guidelines. This presents a unique, empirically validated template for compliant biometric analytics in physical retail.

6.4 Recommendations for Implementation

For SMEs in the early stages of adopting this model, the focus should be on automation and operational stability. The system should act as an unattended background process, with automated ingestion, timestamp synchronisation, and continuous health checks. Real-time inference is discouraged unless stable power, adequate compute, and reliable connectivity are already in place. CCTV installations should be upgraded where necessary to ensure stable angles, appropriate lighting, and consistent image quality. Power continuity, such as through UPS support, is essential in regions affected by load shedding. Batch processing is recommended for demographic inference to reduce costs and avoid reliance on high-performance hardware. Data governance should first adhere to POPIA standards by ensuring transparent communication, rigorous data minimisation, and the deletion of identifiable imagery. Insights need to be conceptualised as decision-support tools to guide product placement, staffing, and promotional timing rather than as real-time operational triggers.

6.5 Limitations

The dataset spans only two trading days from one SME and is therefore neither generalisable nor capable of explaining weekly or seasonal trends. Responses from the survey used to validate this issue were incomplete, leading to uncertainty about ground-truth demographics. Methodologically, only one age-gender model was used, and age estimation remained inconsistent under uncontrolled CCTV conditions. Quality in embedding was also sensitive to lighting variation, overhead angles, and partial occlusions, occasionally fragmenting identity clusters despite post-processing. Environmental factors—such as power instability, camera drift, and timestamp misalignment—also reduced data reliability. This is typical of many SMEs, regardless of size, though it may differ from international retail environments. Based on these considerations, we should regard the findings as suggestive rather than definitive. A wider, multi-site analysis is necessary to establish robustness and transferability.

6.6 Future Research

Improved data quality, model performance, and validation mechanisms should be the priority of future work. Better use of higher-resolution cameras (e.g., 4K) and improved integration of frontal or semi-frontal camera placement would also yield better face crops, reduce embedding noise, and reinforce demographic inference. The pipeline needs to be explored with higher-capacity detection, embedding, and age-estimation models to see how well we can improve accuracy when computational constraints are less stringent. There must be a structured validation process. One such option could be an optional loyalty card scheme where consumers voluntarily submit simple demographic information. That would result in a privacy-compliant validation dataset calibrated to calibrate predictions without infringing on the privacy of the other shoppers. Such studies also need to extend deployment times, adopt multiple cameras, and assess whether operational features run for weeks or months. More studies across a wide variety of SME environments would make it much clearer which sections of the pipeline generalise well and which need modification. Real-world deployments still rely on automation improvements, e.g., ingestion, timestamp alignment, error recovery, and model scheduling. Therefore, a pivotal question that remains is: How can we optimize computer vision models to not only function effectively in resource-constrained environments but also exceed current benchmarks in demographic inference accuracy? This challenge can serve as a catalyst for future explorations focused on obtaining the highest possible performance while maintaining compliance with privacy standards. Taken together, all these developments would allow the present proof-of-concept to evolve into a powerful, scalable framework for supporting SMEs in an evolving data-centric retail environment.

References

Software, APIs, and Models

1. uiprotect. (2024). Unofficial UniFi Protect Python API and CLI (Version 0.6.1) [Computer software]. GitHub. https://github.com/uilibs/uiprotect
2. Ultralytics. (2024). YOLOv11 models. https://docs.ultralytics.com/models/yolo11/
3. Rosebrock, A. (2024). imutils (Version latest). PyPI. https://pypi.org/project/imutils/
4. Tencent ARC. (2022). GFPGAN: Towards real-world blind face restoration. https://github.com/TencentARC/GFPGAN
5. Tencent ARC. (2022). GFPGAN v1.4 release. https://github.com/TencentARC/GFPGAN/releases
6. IvLabs. (2024). FSRCNN model repository. https://github.com/IvLabs/Summer-Projects/tree/main/Summer%202024/FSRCNN
7. HDBSCAN. (2016). Hierarchical density based spatial clustering of applications with noise. https://hdbscan.readthedocs.io/en/latest/
8. NetworkX. (2025). NetworkX: Network analysis in Python. https://networkx.org/
9. Scikit-learn. (n.d.). sklearn.preprocessing.normalize. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.normalize.html
10. nateraw. (2024). vit-age-classifier [Model]. Hugging Face. https://huggingface.co/nateraw/vit-age-classifier
11. prithivMLmods. (2024a). Realistic-Gender-Classification [Model]. Hugging Face. https://huggingface.co/prithivMLmods/Realistic-Gender-Classification
12. prithivMLmods. (2024b). facial-age-detection [Model]. Hugging Face. https://huggingface.co/prithivMLmods/facial-age-detection
13. Fanclan. (2023). Age-gender-model [Pretrained model]. Hugging Face. https://huggingface.co/fanclan/age-gender-model
14. Jocher, G., Chaurasia, A., Qiu, J., & Ultralytics. (2023). YOLO by Ultralytics (YOLOv8) [Software]. https://github.com/ultralytics/ultralytics , https://docs.ultralytics.com/models/yolov8/
15. Jocher, G., Qiu, J., & Ultralytics. (2024). Ultralytics YOLO11 (version 1.0) [Computer software]. Ultralytics Docs. https://docs.ultralytics.com/models/yolo11/
16. Wolf, O. (2025a). 8_hugging_face_models_gender.ipynb [Jupyter Notebook]. GitHub. https://github.com/oskar-wolf/thesis_demographics_cv/blob/main/notebooks/8_hugging_face_models_gender.ipynb

17. Wolf, O. (2025b). 8_hugging_face_models_age.ipynb [Jupyter Notebook]. GitHub. https://github.com/oskar-wolf/thesis_demographics_cv/blob/main/notebooks/8_hugging_face_models_age.ipynb
18. Wolf, O. (2025c). 6_facial_embedding.ipynb [Jupyter Notebook]. GitHub. https://github.com/oskar-wolf/thesis_demographics_cv/blob/main/notebooks/6_facial_embedding.ipynb
19. Wolf, O. (2025d). 7_accounting_allocation.ipynb [Jupyter Notebook]. GitHub. https://github.com/oskar-wolf/thesis_demographics_cv/blob/main/notebooks/7_accounting_allocation.ipynb
20. Wolf, O. (2025e). 9_survey.ipynb [Jupyter Notebook]. GitHub. https://github.com/oskar-wolf/thesis_demographics_cv/blob/main/notebooks/9_survey.ipynb
21. Wolf, O. (2025f). 10_evaluation.ipynb [Jupyter Notebook]. GitHub. https://github.com/oskar-wolf/thesis_demographics_cv/blob/main/notebooks/10_evaluation.ipynb
22. Wolf, O. (2025g). 4_yunet_detect.ipynb [Jupyter Notebook]. GitHub. https://github.com/oskar-wolf/thesis_demographics_cv/blob/main/notebooks/4_yunet_detect.ipynb
23. Wolf, O. (2025h). 5_image_processing.ipynb [Jupyter Notebook]. GitHub. https://github.com/oskar-wolf/thesis_demographics_cv/blob/main/notebooks/5_image_processing.ipynb
24. Wolf, O. (2025i). 3_yolo_person.ipynb [Jupyter Notebook]. GitHub. https://github.com/oskar-wolf/thesis_demographics_cv/blob/main/notebooks/3_yolo_person.ipynb
25. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020). Transformers: State-of-the-art Natural Language Processing. GitHub. https://github.com/huggingface/transformers

Hardware and Devices

26. Ubiquiti Inc. (2023a). Camera G4 Dome. Ubiquiti EU Store. https://eu.store.ui.com/eu/en/products/uvc-g4-dome
27. Ubiquiti Inc. (2023b). UniFi Cloud Key Gen2 Plus and related NVR products. Ubiquiti EU Store. https://eu.store.ui.com/eu/en/category/all-cameras-nvrs
28. Ubiquiti Inc. (2023c). UniFi Dream Router 7 and networking gear. Ubiquiti EU Store. https://eu.store.ui.com/eu/en/category/all-cloud-gateways
29. Security Camera Shop. (2023). 4MP AcuSense Fixed Dome Network Camera New. https://www.securitycamerashop.eu/en/4mp-acusense-fixed-dome-network-camera-new.html

Computer Vision and Face Detection Literature

30. Yu, J., Qi, C. R., & others. (2023). YuNet: A Tiny, Millisecond-level Face Detector. https://www.researchgate.net/publication/370122920_YuNet_A_Tiny_Millisecond-level_Face_Detector
31. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv. https://doi.org/10.48550/arXiv.2004.10934
32. OpenCV CLACHE. (n.d.). CLACHE — Contrast Limited Adaptive Histogram Equalization. https://docs.opencv.org/3.4/d6/db6/classcv_1_1CLAHE.html
33. OpenCV Haarcascade. (n.d). haarcascade_frontalface_alt2.xml. https://github.com/opencv/opencv/blob/master/data/haarcascades/haarcascade_frontalface_alt2.xml
34. OpenCV. Nelon, P. (2022). OpenCV face detection: Cascade classifier vs. YuNet [Blog post]. https://opencv.org/blog/opencv-face-detection-cascade-classifier-vs-yunet/
35. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. . https://doi.org/10.1109/CVPR.2005.177
36. Deng, J., Guo, J., Niannan, X., & Zafeiriou, S. (2020). RetinaFace: Single-shot multi-level face localization in the wild. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5203–5212. https://doi.org/10.1109/CVPR42600.2020.00525
37. Eidinger, E., Enbar, R., & Hassner, T. (2014). Age and gender estimation of unfiltered faces. IEEE Transactions on Information Forensics and Security, 9(12), 2170–2179. https://doi.org/10.1109/TIFS.2014.2359646
38. Viola, P.A., & Jones, M.J. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 1, I-I. https://doi.org/10.1109/CVPR.2001.990517
39. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10), 1499–1503. https://doi.org/10.1109/LSP.2016.2603342
40. Ranjan, R., Sankaranarayanan, S., Castillo, C.D., & Chellappa, R. (2016). An All-In-One Convolutional Neural Network for Face Analysis. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 17-24. https://doi.org/10.1109/FG.2017.137
41. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4685–4694. https://doi.org/10.1109/CVPR.2019.00482

42. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 815–823. https://doi.org/10.1109/CVPR.2015.7298682
43. Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. British Machine Vision Conference (BMVC). https://doi.org/10.5244/C.29.41
44. Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1701–1708. https://doi.org/10.1109/CVPR.2014.220
45. King, D. E. (2009). Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research, 10, 1755–1758. https://doi.org/10.5555/1577069.1755843
46. Serengil, S. T., & Ozpinar, A. (2020). DeepFace: A lightweight face recognition and facial attribute analysis (age, gender, emotion and race) library for Python. GitHub repository. https://github.com/serengil/deepface
47. S. Serengil and A. Ozpinar (2024), "A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules", Journal of Information Technologies, vol. 17, no. 2, pp. 95-107.
48. S. I. Serengil and A. Ozpinar (2020) "LightFace: A Hybrid Deep Face Recognition Framework", 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 23-27.
49. S. I. Serengil and A. Ozpinar (2021) , "HyperExtended LightFace: A Facial Attribute Analysis Framework", 2021 International Conference on Engineering and Emerging Technologies (ICEET), pp. 1-4.
50. Sapkota, R., Flores-Calero, M., Qureshi, R. et al. (2025). YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series. Artif Intell Rev 58, 274. https://doi.org/10.1007/s10462-025-11253-3
51. Tian, X., Ye, Y., & Doermann, D. (2025). YOLOv12: Attention-Centric Real-Time Object Detectors. https://doi.org/10.48550/arXiv.2502.12524

Demographic and Attribute Classification in Faces

52. Khan, K., Attique, M., Khan, R. U., Syed, I., & Chung, T.-S. (2020). A Multi-Task Framework for Facial Attributes Classification through End-to-End Face Parsing and Deep Convolutional Neural Networks. Sensors, 20(2), 328. https://doi.org/10.3390/s20020328
53. Alghaili, M., Li, Z., & Ali, H. A. R. (2020). Deep feature learning for gender classification with covered/camouflaged faces. IET Image Processing, 14(15), 3957–3964. https://doi.org/10.1049/iet-ipr.2020.0199

54. Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 1–15.
55. Guo, G., & Mu, G. (2010). Human age estimation: What is the influence across race and gender? 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops https://doi.org/10.1109/CVPRW.2010.5543609
56. Yang, X., Gao, B., Xing, C., Huo, Z., Wei, X., Zhou, Y., Wu, J., & Geng, X. (2015). Deep Label Distribution Learning for Apparent Age Estimation. 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), 344-350.
57. Jain, A.K., Dass, S.C., & Nandakumar, K. (2004). Soft Biometric Traits for Personal Recognition Systems. International Conference on Biometric Authentication.
58. Karkkainen, K., & Joo, J. (2021). FairFace: Face attribute dataset for balanced race, gender, and age. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 1548–1558.
59. Klare, B. F., et al. (2012). Face recognition performance: Role of demographic information. IEEE Transactions on Information Forensics and Security, 7(6), 1789–1801.
60. Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 34–42.
61. Liu, X., Li, S., Kan, M., Zhang, J., Wu, S., Liu, W., Han, H., Shan, S., & Chen, X. (2015). AgeNet: Deeply Learned Regressor and Classifier for Robust Apparent Age Estimation. 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), 258-266. https://doi.org/10.1109/ICCVW.2015.42
62. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., & Berg, A. C. (2016). SSD: Single shot multibox detector. European Conference on Computer Vision, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
63. Raji, I. D., et al. (2020). Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing. Proceedings of the AAAI/ACM Conference on AI Ethics and Society, 145-151
64. Rothe, R., Timofte, R., & Van Gool, L. (2018). Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision, 126(2–4), 144–157.
65. Rothe, R., Timofte, R., & Van Gool, L. (2015). DEX: Deep expectation of apparent age from a single image. Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW), 10–15.

66. Shi, H., et al. (2023). Face-based age estimation using improved Swin Transformer with attention-based convolution. Frontiers in Neuroscience, 17, https://doi.org/10.3389/fnins.2023.1136934
67. Swaminathan, A., Chaba, M., Sharma, D. & Chaba, Y. (2020) .Gender Classification using Facial Embeddings: A Novel Approach . https://doi.org/10.1016/j.procs.2020.03.342
68. Agbo-Ajala, O., Viriri, S., & Oloko-Oba,M . (2022). Apparent age prediction from faces: A survey of modern approaches. Frontiers in Big Data, 5, 1025806. https://doi.org/10.3389/fdata.2022.1025806
69. Lanitis, A., Taylor, C. J., & Cootes, T. F. (2002). Toward automatic simulation of aging effects on face images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4), 442–455. https://doi.org/10.1109/34.993553
70. Zhang, Z., Song, Y., & Qi, H. (2017). Age Progression/Regression by Conditional Adversarial Autoencoder. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4352-4360. https://doi.org/10.1109/CVPR.2017.463
71. Niu, Z., Zhou, M., Wang, L., Xinbo, G., & Gao, X. (2016). Ordinal regression with multiple output CNN for age estimation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), https://doi.org/10.1109/CVPR.2016.532
72. Kotwal & Marcel. (2024). Mitigating Demographic Bias in Face Recognition via Regularized Score Calibration. IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW). https://doi.org/10.1109/WACVW60836.2024.00125
73. NIST. (2019). Face recognition vendor test (FRVT) ongoing. National Institute of Standards and Technology.

Video Analytics and Surveillance in Retail

74. Beck, A. (2020). Video technologies in retailing: New insights on the use of video and video analytics in retail. ECR Retail Loss Group. https://www.ecrloss.com/research/reviewing-the-use-of-video-technologies-in-retailing/
75. Beck, A. (2024). The utilisation of video analytics in retailing – 2024 edition. ECR Retail Loss Group. https://www.ecrloss.com/research/the-utilisation-of-video-analytics-in-retail/
76. Connell, J., Fan, Q., Gabbur, P., Haas, N., & Pankanti, S. (2013). Retail video analytics: An overview and survey. In Proc. SPIE 8663, Video Surveillance and Transportation Imaging Applications. International Society for Optics and Photonics. https://doi.org/10.1117/12.2008899
77. Pelco by Motorola Solutions. (2023). Retail video analytics: Everything you need to know. https://www.pelco.com/blog/video-analytics-in-retail

78. Senior, A. W., et al. (2007). Video analytics for retail. In IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 423–428). https://www.andrewsenior.com/papers/SeniorRetail2007.pdf
79. ThinkLP. (2023). The early adoption of CCTV in retail: A shift in security practices. https://www.thinklp.com/the-early-adoption-of-cctv-in-retail-a-shift-in-security-practices/
80. Tictag. (2023). AI-powered video analytics: Transforming retail intelligence. https://www.tictag.io/blog/ai-powered-video-analytics-transforming-retail-intelligence
81. DTiQ. (2023). History of security cameras: How did CCTV evolve. https://www.dtiq.com/blog/video-surveillance/history-of-security-cameras
82. Growth Market Reports. (2025). Video analytics for retail market research report 2033 (Base year 2024). https://growthmarketreports.com/report/video-analytics-for-retail-market
83. Hodgson, K. (2025). Challenges & headwinds impact video surveillance tech adoption. SDM Magazine, February 2025. https://www.sdmmag.com/articles/103945-challenges-and-headwinds-impact-video-surveillance-tech-adoption

Retail Analytics, Digital Transformation, and SMEs

84. Admetrics. (2025). Why digital-to-physical retail integration is key to scalable DTC growth. https://www.admetrics.io/en/post/digital-to-physical-retail-integration-dtc-ecommerce-28721
85. Business.com., Farlie, M. (2025). How data analytics impacts small businesses. https://www.business.com/articles/the-state-of-data-analytics/
86. IoT For All, Renno, R. (2019). Retail analytics in physical stores: A path to omni-channel success. https://www.iotforall.com/retail-analytics-omni-channel-success
87. Lumenalta. (2025). 9 use cases of computer vision in retail. https://lumenalta.com/insights/9-use-cases-of-computer-vision-in-retail
88. LS Retail, Jonnson, A.. (2024). How RFID inventory systems can improve stock management in stores. https://www.lsretail.com/resources/how-rfid-inventory-systems-can-improve-stock-management-in-stores
89. MIT Sloan Management Review, Chatterjee ,S.C. (2025). The future of physical retail: 5 actions to elevate customer experience. https://mitsloan.mit.edu/ideas-made-to-matter/future-physical-retail-5-actions-to-elevate-customer-experience
90. RetailNext, Kirsten, A. (2023). 3 in-store analytics case studies every retailer should read. https://retailnext.net/blog/3-in-store-analytics-case-studies-every-retailer-should-read
91. RudderStack , Varangaonkar, A. (2021). What is clickstream analytics? A gamechanger for e-commerce. https://www.rudderstack.com/blog/clickstream-analytics-gamechanger-for-ecommerce/

92. Sakovich, N. (2020). IoT in retail and e-commerce: Effective application scenarios. SaM Solutions Blog. https://sam-solutions.com/blog/iot-in-retail-ecommerce/
93. Zensors. (2023). Why retail stores are lagging behind in digital transformation. https://www.zensors.com/post/why-retail-stores-are-lagging-behind-in-digital-transformation-what-stores-can-do
94. Enefer, S. (2023, July 30). Small data, big advantage: How SMEs can win the analytics race. Clarity Data Consulting Blog. https://www.claritydataconsulting.com/post/small-data-big-advantage-how-smes-can-win-the-analytics-race
95. Kgakatsi, M., Galeboe, O. P., Molelekwa, K. K., & Thango, B. A. (2024). The impact of big data on SME performance: A systematic review. Businesses, 4(4), 632–695. https://doi.org/10.3390/businesses4040038
96. Long, T. (2025, August 27). How digital services empower SMEs and start-ups. Information Technology & Innovation Foundation. https://itif.org/publications/2025/08/27/how-digital-services-empower-smes-and-start-ups/
97. OECD. (2021). The digital transformation of SMEs. OECD Publishing. https://www.oecd.org/en/publications/the-digital-transformation-of-smes_bdb9256a-en.html
98. Patterson-Waites, A. (2023, July 14). Smaller and mid-sized businesses are fighting for survival. This is how they could prosper. World Economic Forum. https://www.weforum.org/stories/2023/07/digital-transformation-potential-smes/
99. World Economic Forum, & Bainchini,M., Michalkova, V. (2019). Data analytics in SMEs: Trends and policies (OECD Working Papers). OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/06/data-analytics-in-smes_1535d46b/1de6c6a7-en.pdf

Data Protection, Privacy, and Ethics

100. Data Protection Laws of the World. (2025). Data protection laws in South Africa. Retrieved from https://www.dlapiperdataprotection.com/index.html?t=law&c=ZA
101. General Data Protection Regulation, Regulation (EU) 2016/679, Arts. 4(14), 6, 9, 17 (2016).
102. Information Commissioner's Office. (2022). Facial Recognition Technology (FRT) and surveillance. ICO Guidance on Video Surveillance. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/cctv-and-video-surveillance/guidance-on-video-surveillance-including-cctv/additional-considerations-for-technologies-other-than-cctv/facial-recognition-technology-frt-and-surveillance/
103. Keylabs. (2024). The ethical gaze: Examining bias and privacy concerns in AI image recognition. Keylabs Blog. https://keylabs.ai/blog/the-ethical-gaze-examining-bias-and-privacy-concerns-in-ai-image-recognition/

104. Lewis Silkin LLP, Laher, Z. . (2025). Surveillance or safeguard? The CNIL's verdict on augmented cameras for age verification. https://www.lewissilkin.com/en/insights/2025/07/25/surveillance-or-safeguard-the-cnils-verdict-on-augmented-cameras-for-age-verifi-102kxrf

105. Michalsons. (2024). Biometric laws around the world. Michalsons Blog. https://www.michalsons.com/blog/biometrics-laws-around-the-world/42094

106. Protection of Personal Information Act 4 of 2013 (POPIA). (2013). South Africa. https://www.justice.gov.za/legislation/acts/2013-004.pdf

107. Reuters. (2020, November 26). French watchdog fines Carrefour 3 mln euros for privacy rule breaches. https://www.reuters.com/business/french-watchdog-fines-carrefour-3-mln-euros-privacy-rule-breaches-2020-11-26/

108. Tencent Cloud. (2024). How to build a GDPR-compliant facial recognition system? Tencent Cloud Techpedia. https://www.tencentcloud.com/techpedia/120225

109. Xiang, A. (2022). Being "seen" versus "mis-seen": Tensions between privacy and fairness in computer vision. Harvard Journal of Law & Technology, 36, 1–65. https://jolt.law.harvard.edu/assets/articlePDFs/v36/Xiang-Being-Seen-Versus-Mis-Seen.pdf

Cloud Computing and Pricing

110. Amazon Web Services. (2023a). Amazon S3 pricing. https://aws.amazon.com/s3/pricing/

111. Amazon Web Services. (2023b). Understanding AWS pricing for EC2, S3, EBS, RDS, & more. https://www.finout.io/blog/understanding-aws-pricing

112. Nutanix. (2023). 3 ways cloud computing powers economies of scale. https://www.nutanix.com/how-to/3-ways-cloud-computing-powers-economies-of-scale

Appendices

Appendix A — Project Code Resources and File Structure

The full project is available on GitHub at:

https://github.com/oskar-wolf/thesis_demographics_cv

The repository is structured into modular directories corresponding to each stage of the analytical pipeline. This setup ensures reproducibility and traceability from raw CCTV footage to final analytical outputs. To comply with POPIA-aligned data minimisation practices, raw face images, full-resolution frames, and any identifiable intermediate outputs were not uploaded to GitHub. These are excluded through the project's .gitignore configuration. All sensitive data was automatically deleted following processing. Only anonymised facial embeddings and a limited set of non-identifiable cropped examples used for evaluation were retained locally and referenced in this thesis.

All execution environments were established using Conda .yml files. Additional packages were installed with pip as needed. For reproducibility, Conda dependencies are stored in environment/<env>_environment.yml. Pip-installed libraries are listed in environment/<env>_pip_requirements.txt. This structure enables full cross-platform recreation of the exact runtime used throughout the project.

Folder	Description
data/	Raw CCTV footage, extracted frames, face crops, and all prediction outputs.
db/	Exported POS datasets and merged transaction files.
environment/	Conda .yml environment files containing all dependencies for reproducibility.
models/	YOLO weights, Haar Cascades, FSRCNN, and GFPGAN models used for inference.
notebooks/	End-to-end pipeline Jupyter notebooks (01–10).
results/	All generated analytics, figures, datasets, and evaluation outputs.
EDA/	Additional exploratory analysis and JSON summaries.
evaluation/	Combined model comparison outputs and aggregated performance summaries.
survey/	Raw customer survey responses.
thesis_docs/	Folder containing written thesis

Table A1: Project Folder Structure

Notebook	Environment	Purpose
1_video_download.ipynb	unifi_cv	Downloading CCTV footage and extracting raw video streams from UniFi
2_db_connect.ipynb	cv_38	Loading, cleaning, and merging POS datasets
3_yolo_person.ipynb	cv_38	Person detection using YOLO models
4_yunet_detect.ipynb	cv_38	Face detection using YuNet (OpenVINO)
5_image_processing.ipynb	face_enhance	Face enhancement: FSRCNN, GFPGAN, preprocessing
6_facial_embedding.ipynb	cv_38	Embedding extraction and identity clustering
7_accounting_allocation.ipynb	cv_38	Timestamp matching, aligning faces with POS transactions
8_hugging_face_models_age.ipynb	hugging_face	Transformer-based age estimation
8_hugging_face_models_gender.ipynb	hugging_face	Transformer-based gender estimation
9_survey.ipynb	cv_38	Survey import, preparation, and comparison with model outputs
10_evaluation.ipynb	cv_38	Full analytical evaluation: accuracy, temporal patterns, revenue by demographic

Table A2: Conda Environment for Notebooks

Appendix B — System Architecture and Computational Environment

Component	Specification
CPU	AMD Ryzen 7 2700X Eight-Core (~3.7 GHz, 16 logical threads)
GPU	NVIDIA GeForce RTX 2060 (5.9 GB VRAM)
RAM	64 GB
Operating System	Windows 11 Home (64-bit)
Storage	Approx. 135 GB SSD allocated for raw footage
Generated Files	Approx. 147,000 files in 1,080 folders during batch inference

Table B1: System Specifications

Appendix C — Final Dataset Schema and Extended Model Evaluation

Column Name	Data Type	Description
image_name	str	Filename of the clustered face image linked to POS
cluster_id	int	Cluster index for facial identity grouping
AccoID	int	Unique ID from accounting record
AccoDocNo	str	Document reference number
AccoDate	datetime	Original POS transaction datetime
timestamp	datetime	Timestamp from camera match
AccoAmount	float	Transaction amount
Quantity	int	Quantity of product purchased
Discount	float	Discount applied
UnitPrice	float	Price per unit
StockCateDesc	str	Product category description
StockName	str	Product name
StockDescription	str	Product detailed description
predicted_gender	str	Model-predicted gender
confidence	float	Gender prediction confidence
predicted_age	str	Predicted age group (e.g., "31–40")
age_confidence	float	Confidence of age group prediction
InvNo	int	Invoice or transaction grouping number
survey_gender	str	Gender self-reported in optional survey
survey_age_group	str	Age group self-reported in survey
survey_time	str	Time when survey was completed
survey_completed	bool / str	Flag indicating if survey was submitted ("TRUE"/"FALSE")

Table C1: Final Dataset Schema After Preprocessing

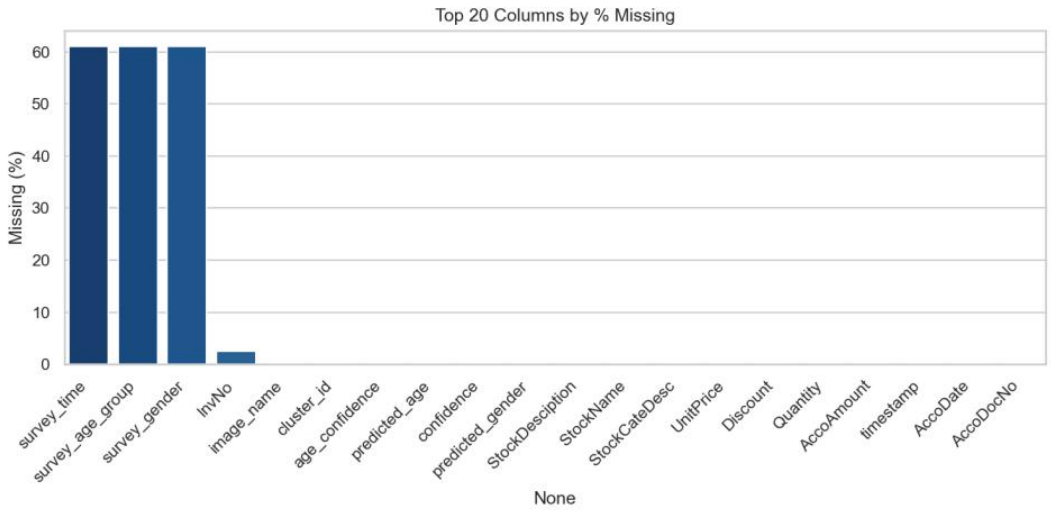

Figure C2: Top 20 Columns by % Missing

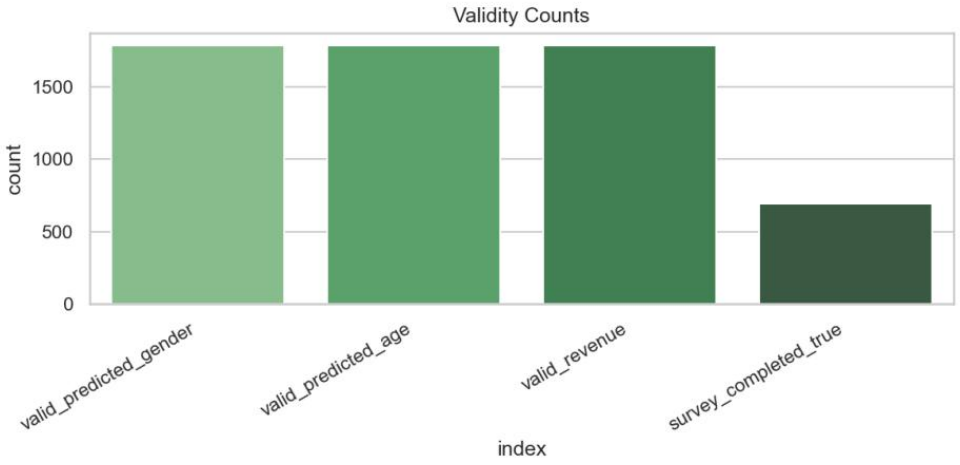

Figure C3: Data Validity Summary

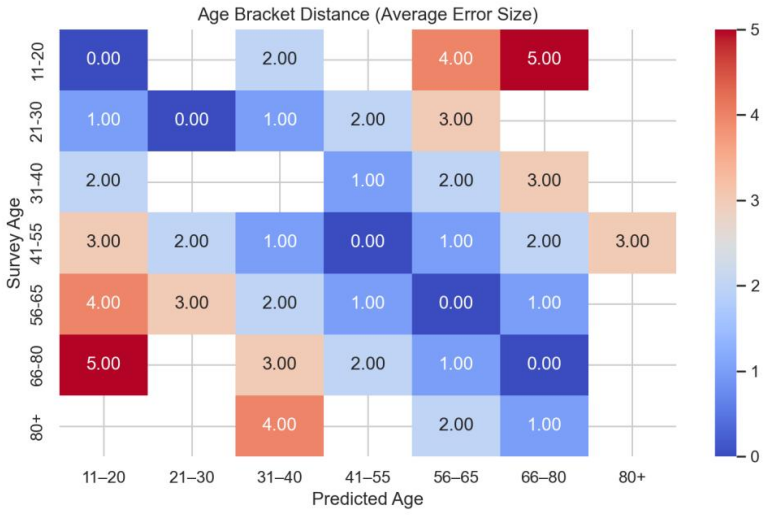

Figure C4: Age Bracket Distance Heatmap

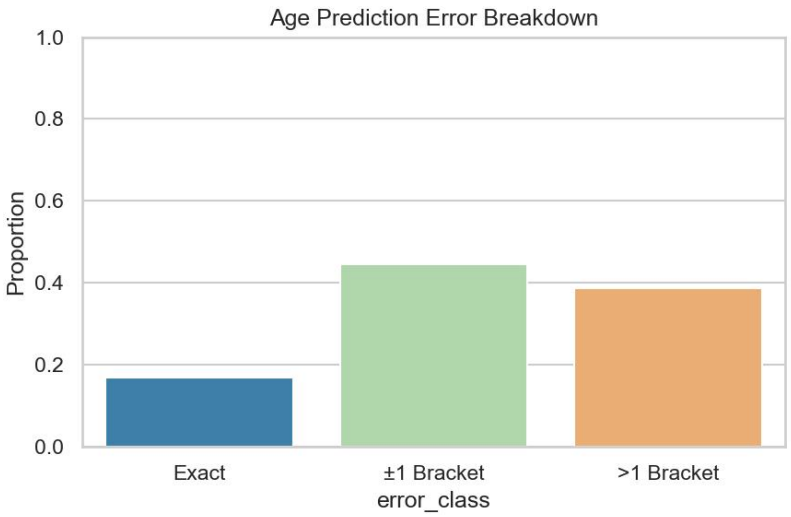

Figure C5: Age Prediction Error Breakdown

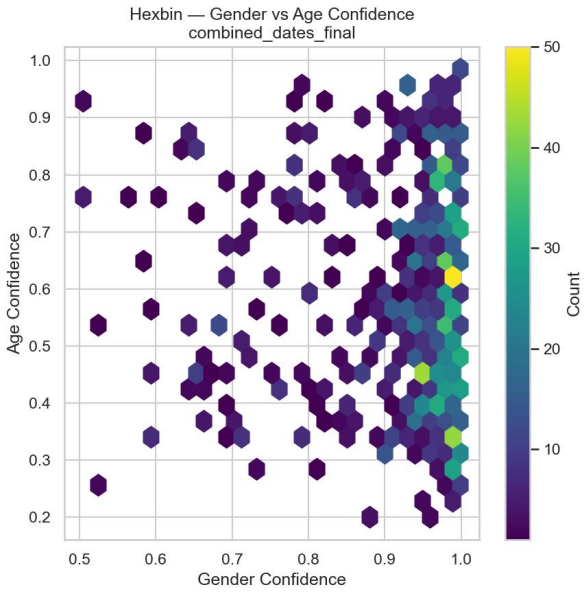

Figure 22: Hexbin Joint Age and Gender Confidence Distribution

Appendix D — Extended Revenue, Category, and Temporal Analytics

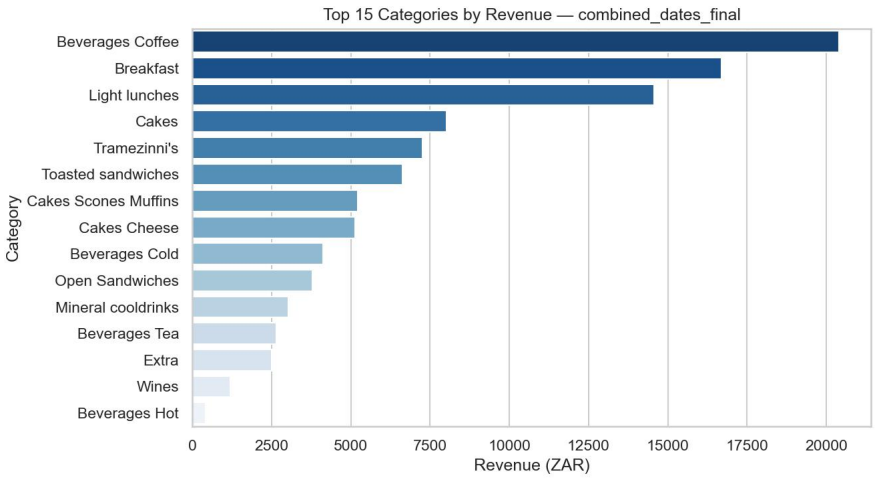

Figure D1: Top 15 Categories by Revenue

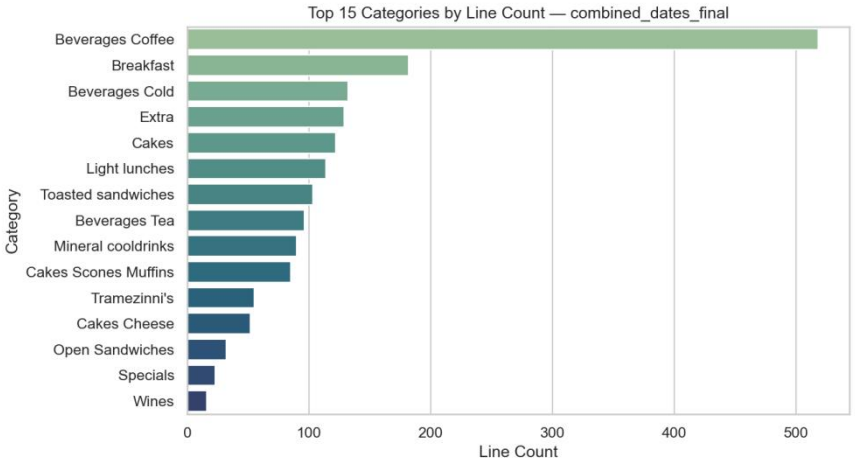

Figure D2: Top 15 Categories by Line Count

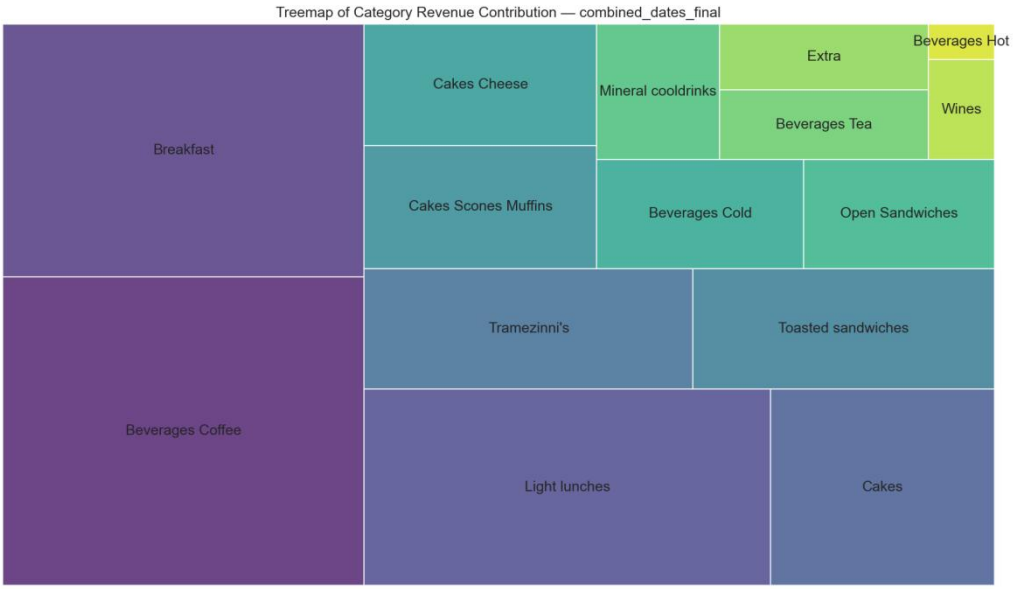

Figure D3:Heatmap of Category Revenue Contribution

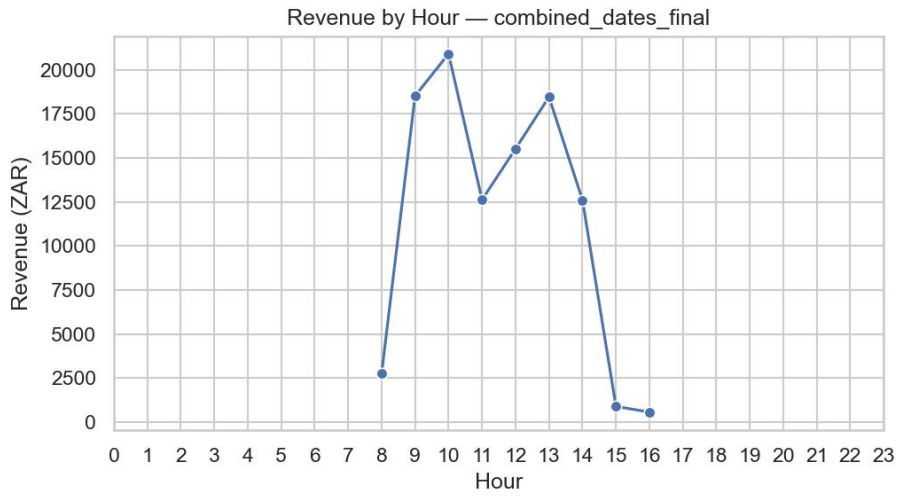

Figure D4: Revenue by Hour

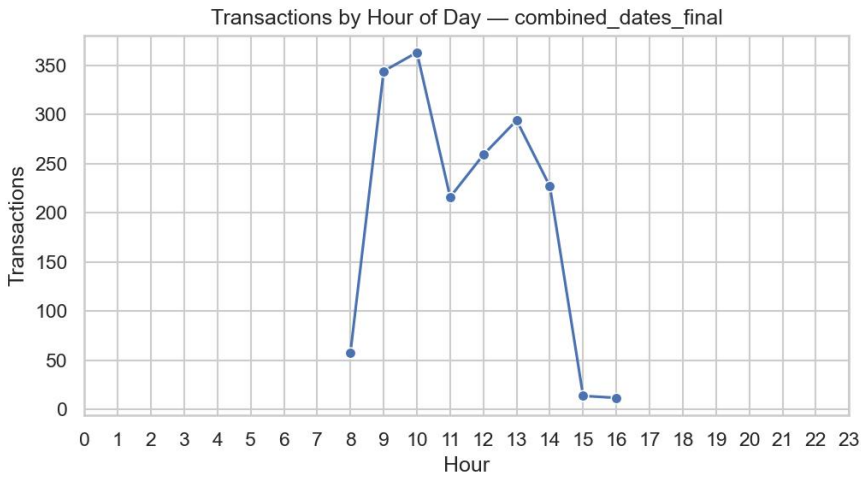

Figure D5: Transactions by Hour

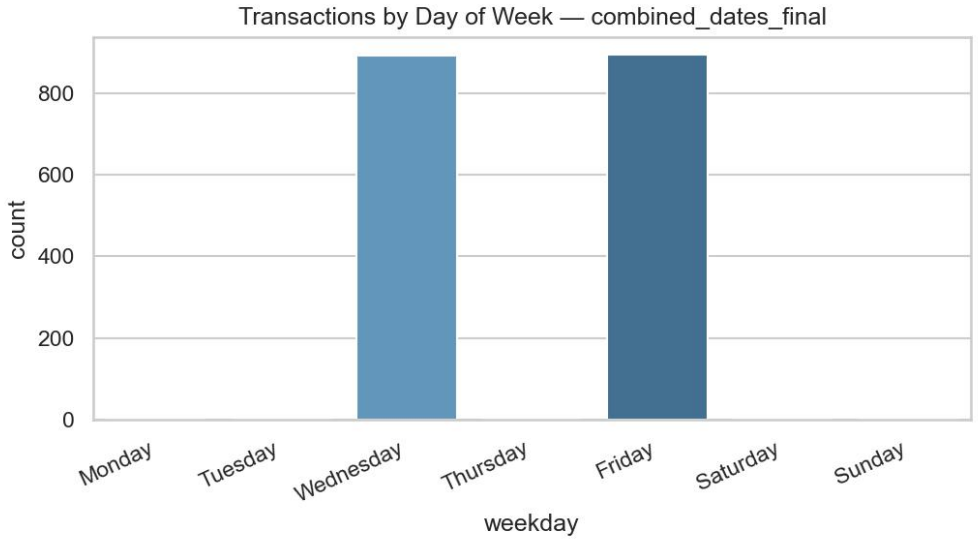

Figure D6: Transactions by Day of the Week

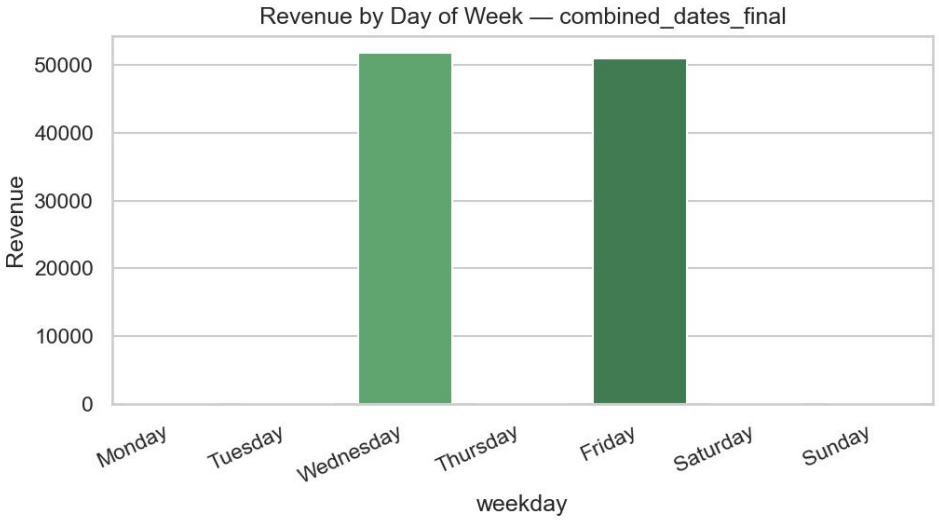

Figure D7: Revenue by Day of the Week

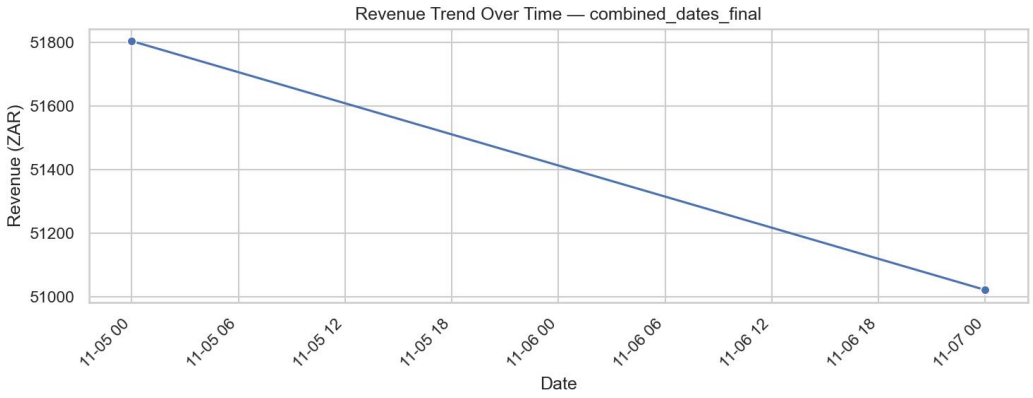

Figure D8: Revenue Trend Over Time

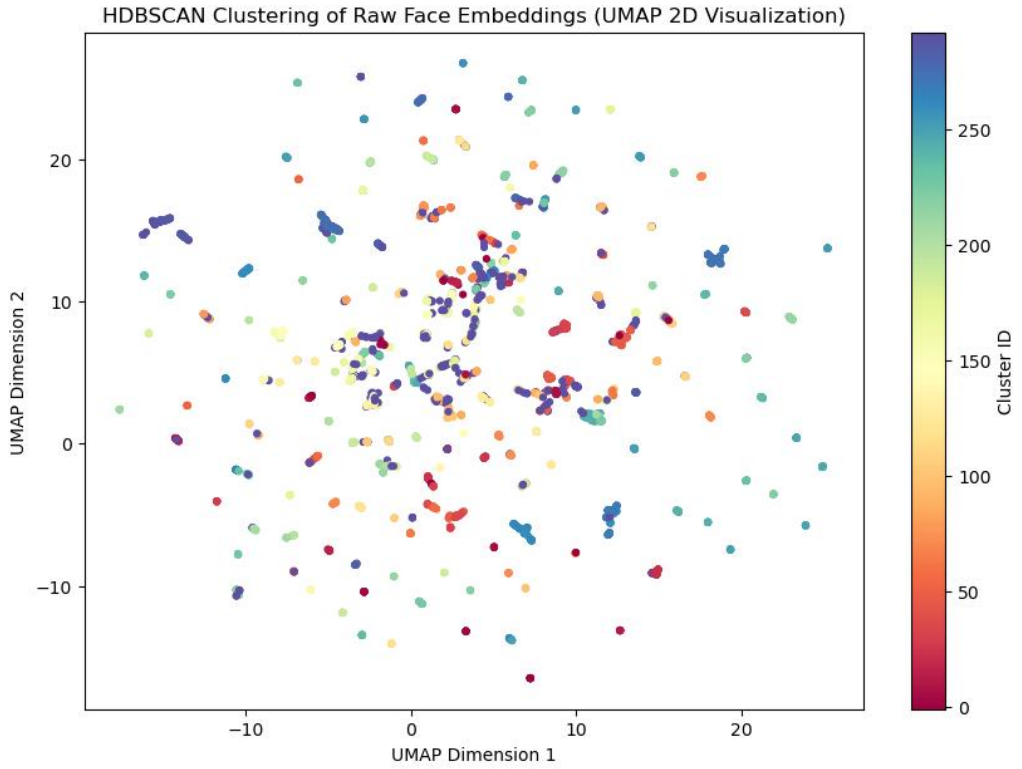

Figure D9: UMAP Scatterplot on HDBSCAN Cluster Embeddings

Appendix E — Raw Survey Materials

The following text appeared on the customer survey form distributed at the point of sale:

“Customer Survey for Academic Research
Purpose: This survey aims to gather information on customer demographics for academic research purposes.

Confidentiality: All responses are anonymous and will be used only for academic analysis.

Participation: Your participation is voluntary and appreciated.”

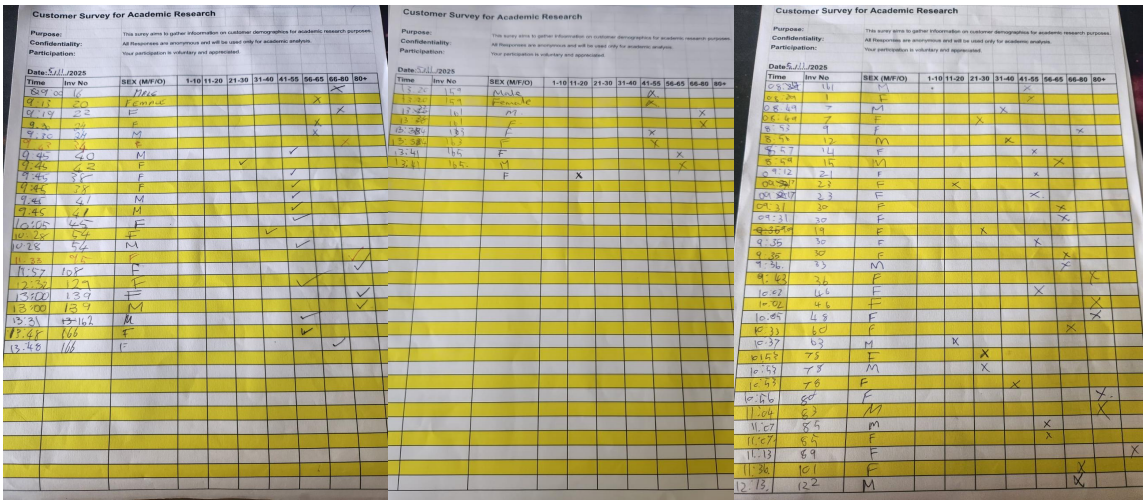

Figure E1: Academic Research Survey 2025-11-5

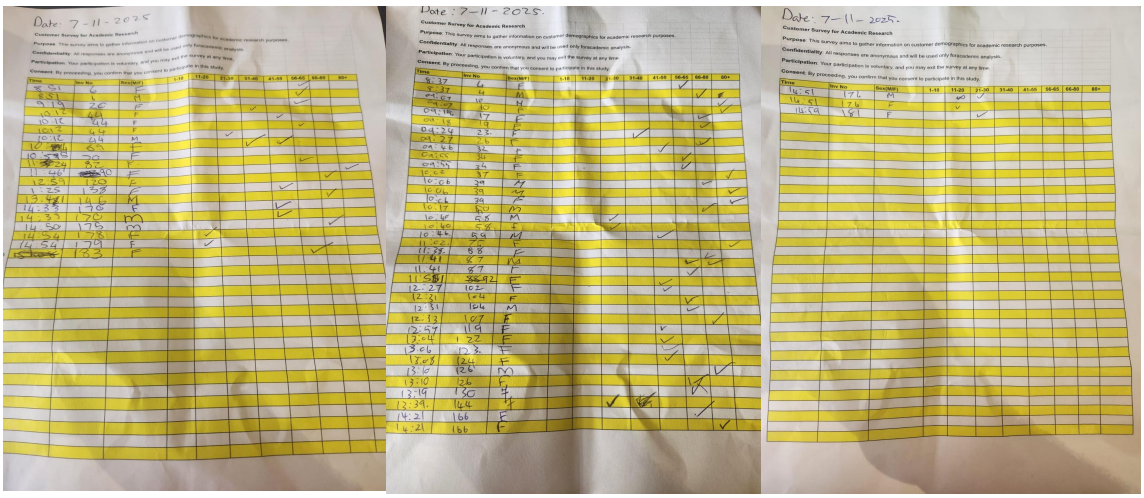

Figure E2: Academic Research Survey 2025-11-7

Participants were asked to provide two demographic details: gender and age group, selected from predefined categories. To ensure accurate alignment between survey responses and point-of-sale data, the cashier recorded the transaction invoice number and the precise transaction time on each survey form at the time of purchase. No names or personal identifiers were collected, and the survey remained fully anonymous in accordance with POPIA-compliant data minimisation practices. The collected survey responses constituted the ground-truth dataset for validating the demographic predictions produced by the computer vision pipeline described in Chapters 3 and 4.

Appendix F — Image Processing and Inference Examples

Figure F1 presents the complete image-processing and inference pipeline employed in this study, demonstrating the transformation of a raw CCTV frame into a demographic prediction. The process initiates with unprocessed point-of-sale (POS)-facing camera footage, from which YOLO identifies the active customer and extracts an individual person crop. YuNet subsequently isolates the facial region to generate a clean face crop suitable for analysis. To enhance image quality under typical small and medium enterprise (SME) CCTV conditions, FSRCNN is applied for super-resolution and GFPGAN for facial restoration. The refined image is then processed by transformer-based models that produce gender and age predictions with associated confidence scores. The example depicted uses images of the researcher, included with full consent, and represents the exact sequence applied to all customer detections during the study. All identifiable images were deleted after processing, with only anonymised embeddings retained in accordance with POPIA-aligned data minimisation.

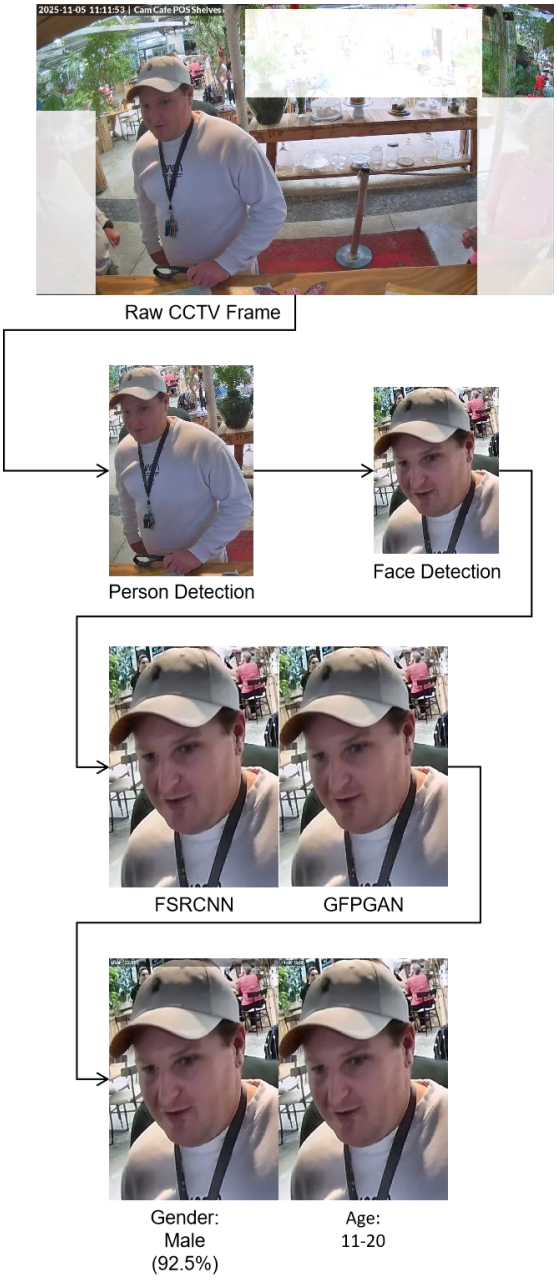

Figure F1: End-to-End CCTV Image Processing and Demographic Inference Pipeline

Appendix G — Declaration of Authenticity

I hereby declare that I have completed this Bachelors/ Master's thesis on my own and without any additional external assistance. I have made use of only those sources and aids specified and I have listed all the sources from which I have extracted text and content. This thesis or parts thereof have never been presented to another examination board. I agree to a plagiarism check of my thesis via a plagiarism detection service.

Place, Date: 2025-12-4

Student Signature: