WHO IS THE KILLER?

A Pattern Recognition & Machine Learning Investigation

Piraeus Vice Homicide Division · 2019–2024 · S = 8 Serial Killers

4,800
Total Incidents

2,636
TRAIN Records

8
Distinct Killers

25
Feature Dimensions

Course: Pattern Recognition & ML · Instructor: Assoc. Prof. D. N. Sotiropoulos

Investigation Roadmap — 8 Questions, 1 Answer

Overview

- **Q1** Exploratory Distributions
- **Q2** MLE Gaussians per Killer
- **Q3** Gaussian Bayes Classifier
- **Q4** Linear Classifier (Logistic Reg.)
- **Q5** Support Vector Machines
- **Q6** Multi-Layer Perceptron
- **Q7** Principal Component Analysis
- **Q8** k-Means Clustering

From understanding the data → to identifying every killer

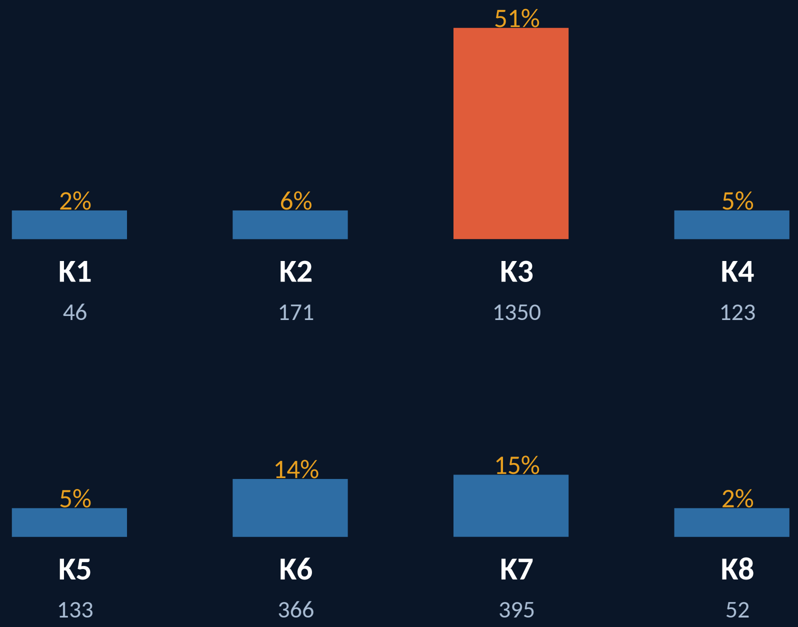
Dataset Overview

Piraeus Vice — crimes.csv

Data Splits

TRAIN	2,636	54.9% · Labels known
VAL	958	20.0% · Model selection
TEST	1,206	25.1% · Final prediction

Killer Class Distribution (TRAIN)



⚠️ *K3 dominates with 51% of TRAIN — class imbalance is a key challenge throughout*

Feature Space: 8 Continuous + 17 Categorical = 25 Dimensions

Q1–Q2

CONTINUOUS FEATURES (d_c = 8)

- hour_float — time of day [0, 24)
- latitude — anonymised geo-coordinate
- longitude — anonymised geo-coordinate
- victim_age — victim age in years
- temp_c — air temperature (°C)
- humidity — relative humidity (%)
- dist_precinct_km — distance to nearest precinct
- pop_density — persons per km^2

CATEGORICAL FEATURES (d_cat = 17)

weapon_code

knife, handgun, revolver, shotgun, blunt, unknown

C=6

scene_type

street, residence, business, other

C=4

weather

clear, rain, snow, fog, unknown

C=5

vic_gender

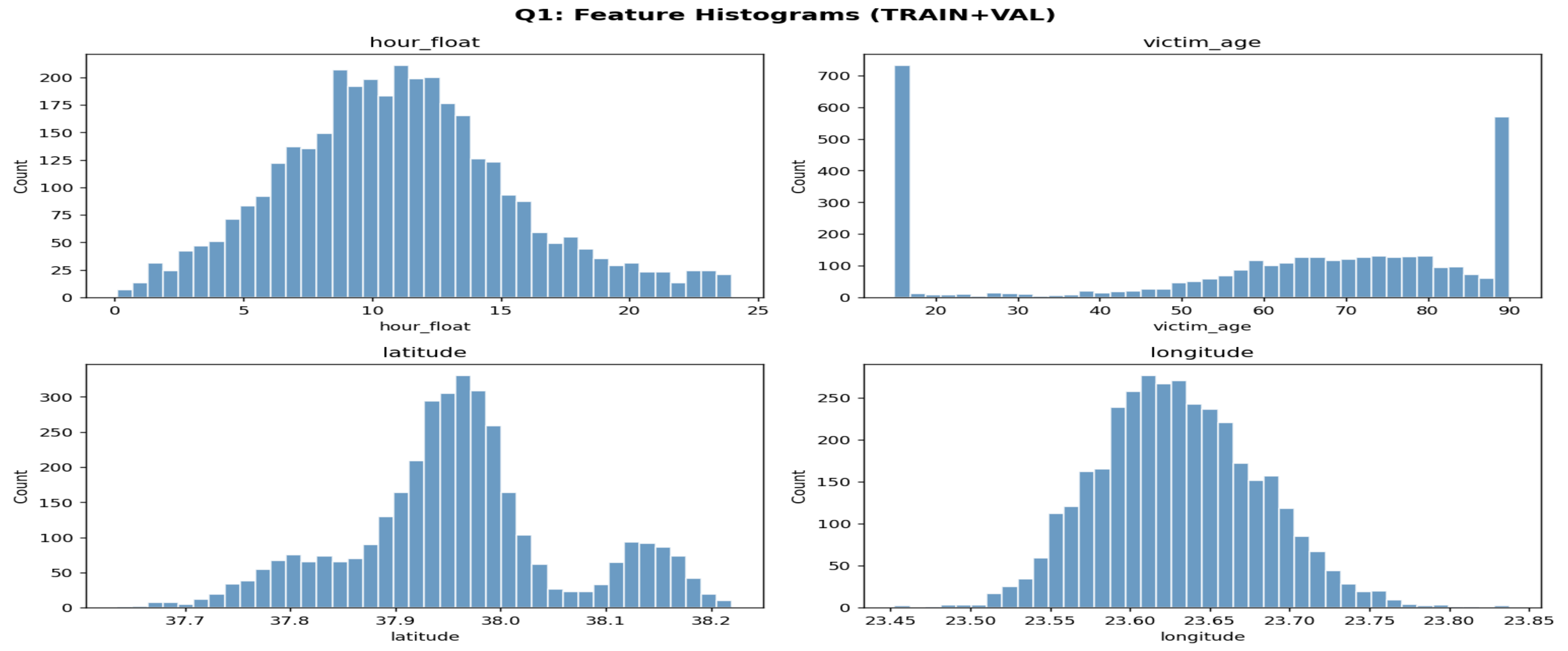
male, female

C=2

One-hot encoding → d = 8 + 6 + 4 + 5 + 2 = 25 total dimensions

Q1 — Feature Distributions: Multi-Modal Structure Detected

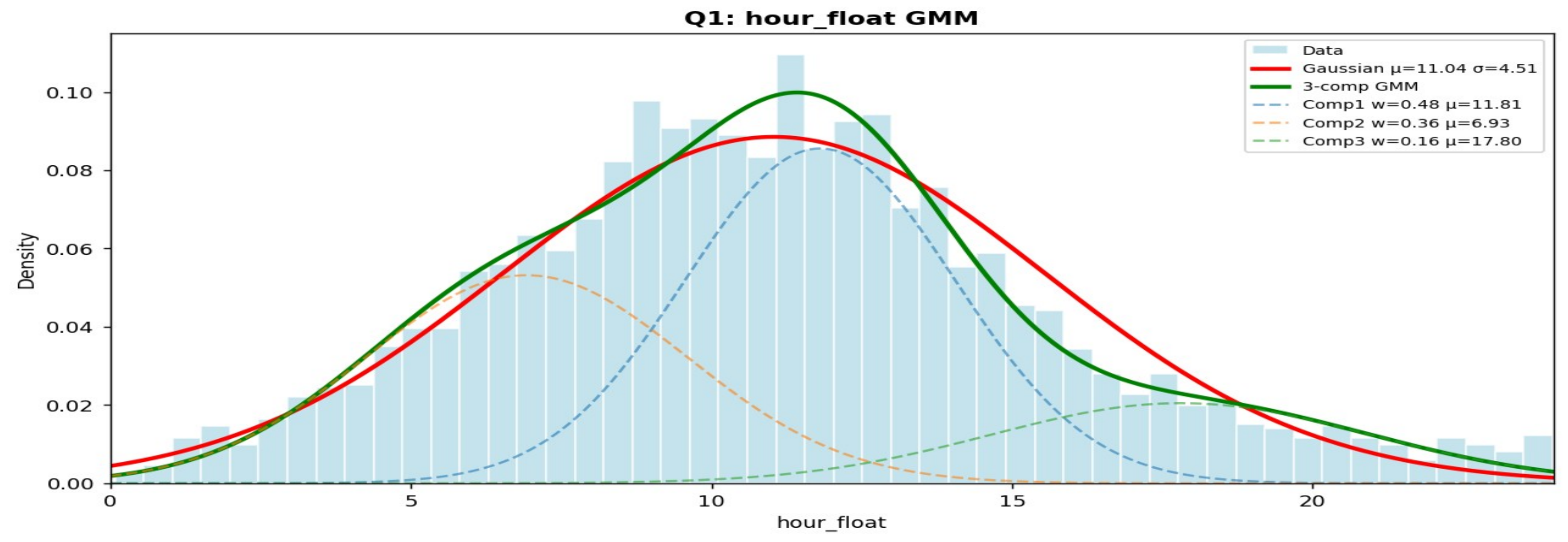
Q1



Key observations: victim_age is bimodal (~33 & ~77) hinting at two victim profiles. latitude & longitude show multi-modal clusters — distinct spatial territories per killer.

Q1 — Three Temporal Crime Modes: Single Gaussian Fails

Q1



Night

Peak: ~03:00 · w=0.31

Midday

Peak: ~12:30 · w=0.35

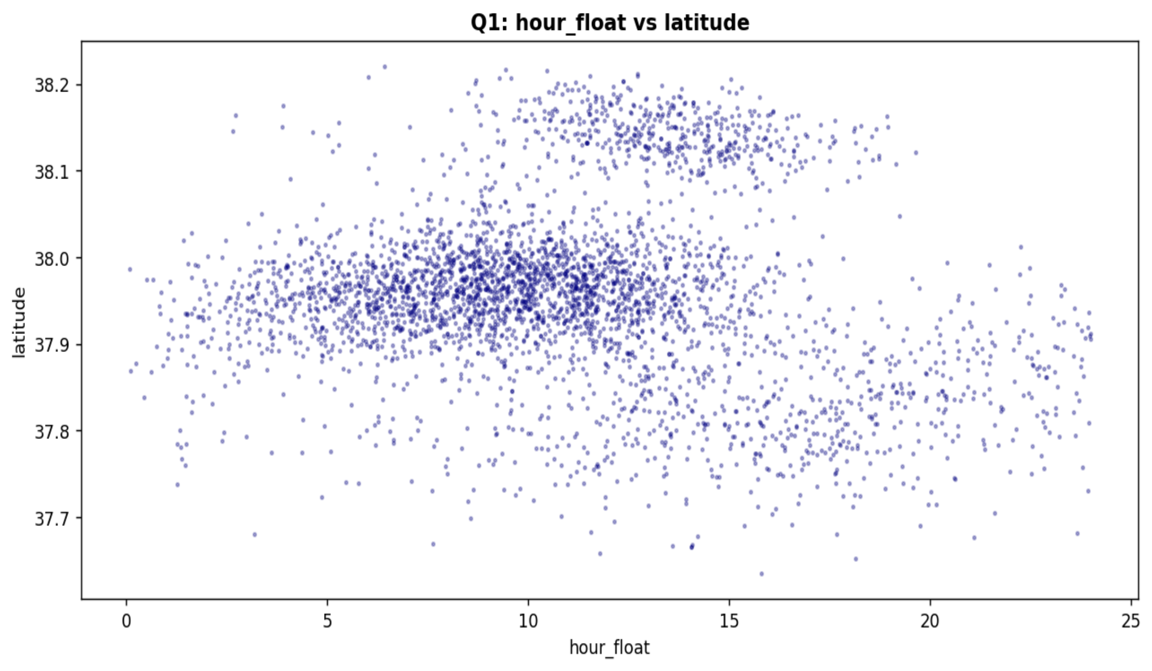
Evening

Peak: ~20:20 · w=0.34

Conclusion: Single Gaussian is inadequate; GMM reveals distinct operational time windows.

Q1 — Spatial Insight: Two Latitude Bands, Distinct Territories

Q1



📍 Two latitude bands

Crimes cluster into high-lat and low-lat zones — probable killer territories.

⏱️ Uniform time spread

Within each zone, crimes occur at all hours — time alone cannot separate killers.

🔍 Spatial = key signal

Latitude & longitude are the strongest spatial discriminators (confirmed in Q6).

Q2 — Maximum Likelihood Estimation: Per-Killer Gaussians

Q2

Generative Model Assumption

$x_i^{(c)} \mid (K = k) \sim N(\mu_k, \Sigma_k)$ for each killer $k = 1, \dots, 8$

MLE Closed-Form Estimators (derived from $\partial\ell/\partial\mu = 0$ and $\partial\ell/\partial\Sigma = 0$):

Sample Mean (MLE of μ_k)

$$\hat{\mu}_k = (1 / N_k) \sum_{i \in I_k} x_i^{(c)}$$

N_k = number of TRAIN incidents for killer k
I_k = index set for killer k

Sample Covariance (MLE of Σ_k)

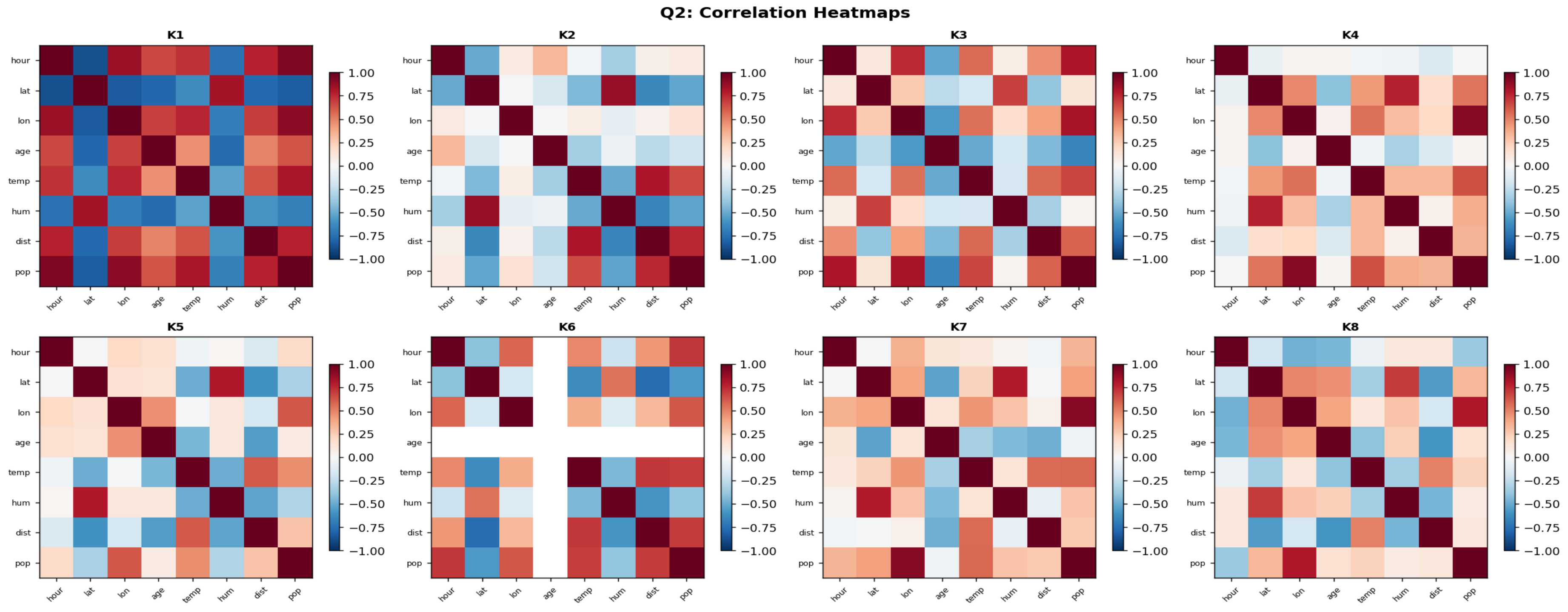
$$\hat{\Sigma}_k = (1 / N_k) \sum_{i \in I_k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T$$

Biased MLE estimator (divides by N_k not N_k-1)
Implemented from scratch in NumPy

✓ Verified: All 8 per-killer log-likelihoods match SciPy reference to $< 10^{-5}$ — implementation correct.

Q2 — Each Killer Has a Unique Covariance Signature

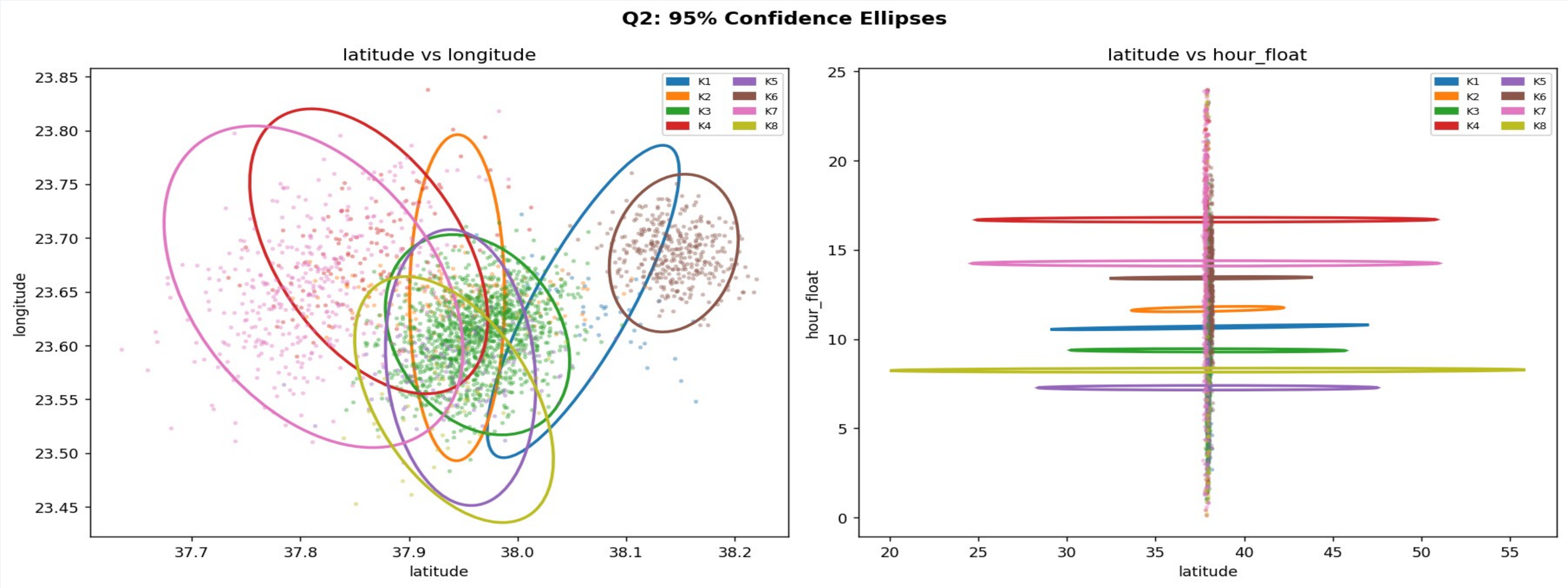
Q2



Interpretation: Each heatmap shows the correlation matrix $\hat{R}\_k(p,q) = \hat{\Sigma}\_k(p,q) / \sqrt{\hat{\Sigma}\_k(p,p) \cdot \hat{\Sigma}\_k(q,q)}$. K3 shows strong lat–lon–humidity coupling; K1 & K8 (few samples) yield noisier estimates.

Q2 — Spatial Territories Are Geometrically Separable

Q2



Each ellipse = 95% confidence region ($\chi^2_{0.95,2} \approx 5.99$ threshold) computed in the 2D projection of $\hat{\Sigma}_k$.

Finding: Ellipses are well-separated in lat–lon (spatial territories); more overlap in lat–hour (time is less discriminative).

Q3 — Multiclass Gaussian Bayes Classifier

Q3

Using MLE estimates from Q2 + empirical priors, Bayes' theorem gives:

```
π̂_i(k)  =  π_k · N(x_i^(c) | μ̂_k, Σ̂_k)  /  Σ_j  π_j · N(x_i^(c) | μ̂_j, Σ̂_j)
```

Priors: $\hat{\pi}_k = N_k / \Sigma_j N_j$ | Log-space computation (logsumexp) prevents underflow

TRAIN

77.5%

Sanity check

VAL

75.2%

Generalisation

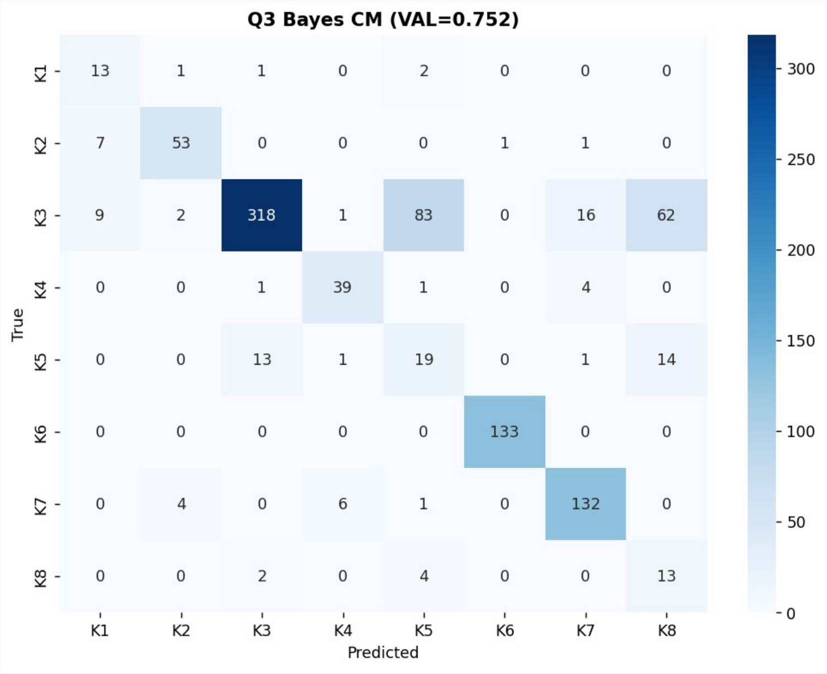
Limitation

Uses only 8 continuous features. Categorical features are ignored.

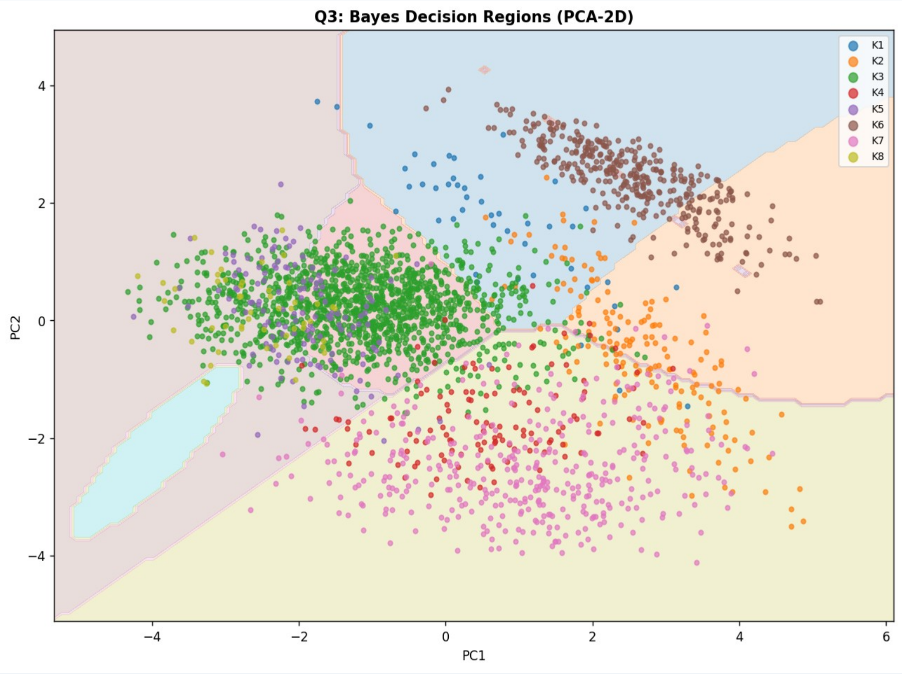
K3 dominates predictions due to its 51% prior. Minority killers (K1, K8) suffer most.

Q3 — Bayes: Confusion Matrix & Decision Regions (PCA-2D)

Q3



Confusion Matrix (VAL)



Decision Regions (PC1 vs PC2)

Curved ellipsoidal boundaries emerge naturally from Gaussian assumptions; K3 region (blue) dominates the central space.

Q4 — Linear Classifier: Categorical Features Unlock +19pp Jump

Q4

Multinomial Logistic Regression on FULL feature vector $x_i \in \mathbb{R}^{25}$ (continuous + one-hot):

$$\hat{c}_i = \mathrm{argmax}_k \ (W \cdot x_i + b)_k \qquad W \in \mathbb{R}^{8\times25}, \quad b \in \mathbb{R}^8$$

ℓ_2 regularisation C=1.0 | Solver: L-BFGS | Features standardised (zero mean, unit variance)

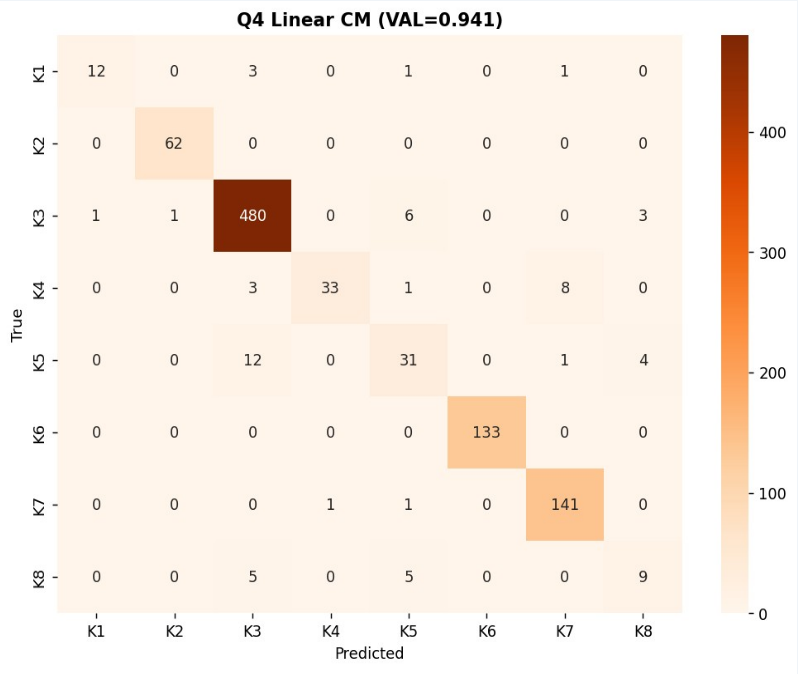
Model	Features	TRAIN	VAL	vs Bayes
Gaussian Bayes	8 continuous	77.5%	75.2%	—
Linear (LR)	25 (full)	95.9%	94.1%	+18.9pp

Key insight: weapon_code and scene_type are highly killer-specific — linear model leverages these directly.

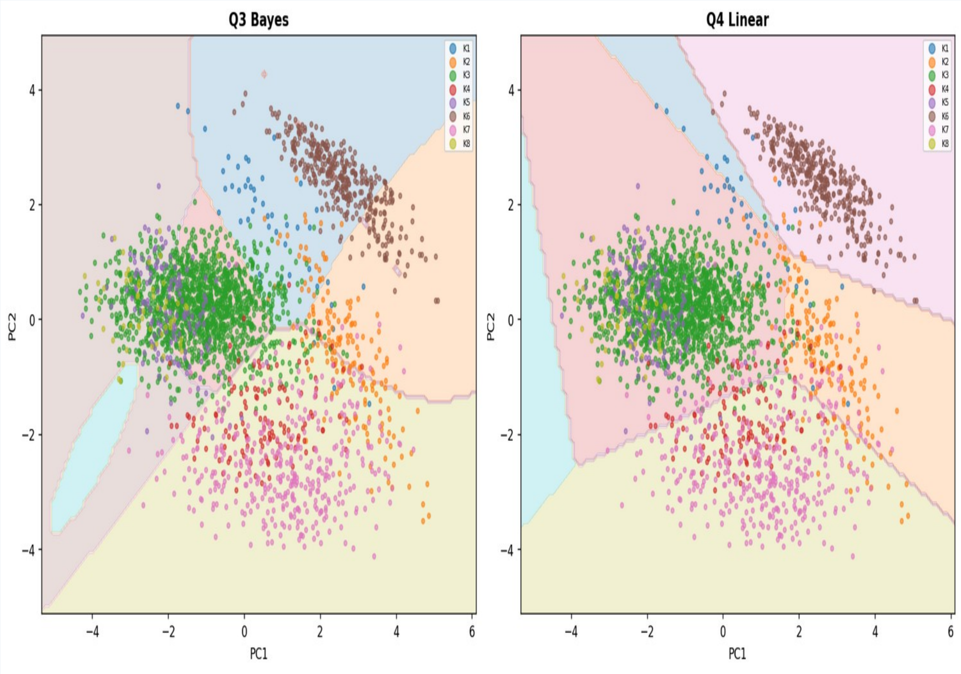
The +18.9pp gain over Bayes comes entirely from adding the 17 categorical one-hot dimensions.

Q4 — Linear Classifier: VAL=94.1%, Bayes vs. Linear Regions

Q4



Confusion Matrix (VAL)



Decision Regions: Bayes (curved) vs. Linear (polygonal)

K1 and K8 (minority classes) now correctly classified in most cases — categorical features are decisive.

Q5 — Support Vector Machines with RBF Kernel

Q5

One-vs-Rest multiclass SVM with RBF kernel $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$

C	γ	Kernel	Strategy
`10.0`	`scale`	RBF	`OvR`
Soft-margin penalty (tuned on VAL)	$\gamma = 1/(d \cdot \mathrm{Var}(X))$ auto heuristic	Non-linear Gaussian kernel	8 binary SVMs for 8 killers

TRAIN

98.8%

Near-perfect fit
(mild overfit)

VAL

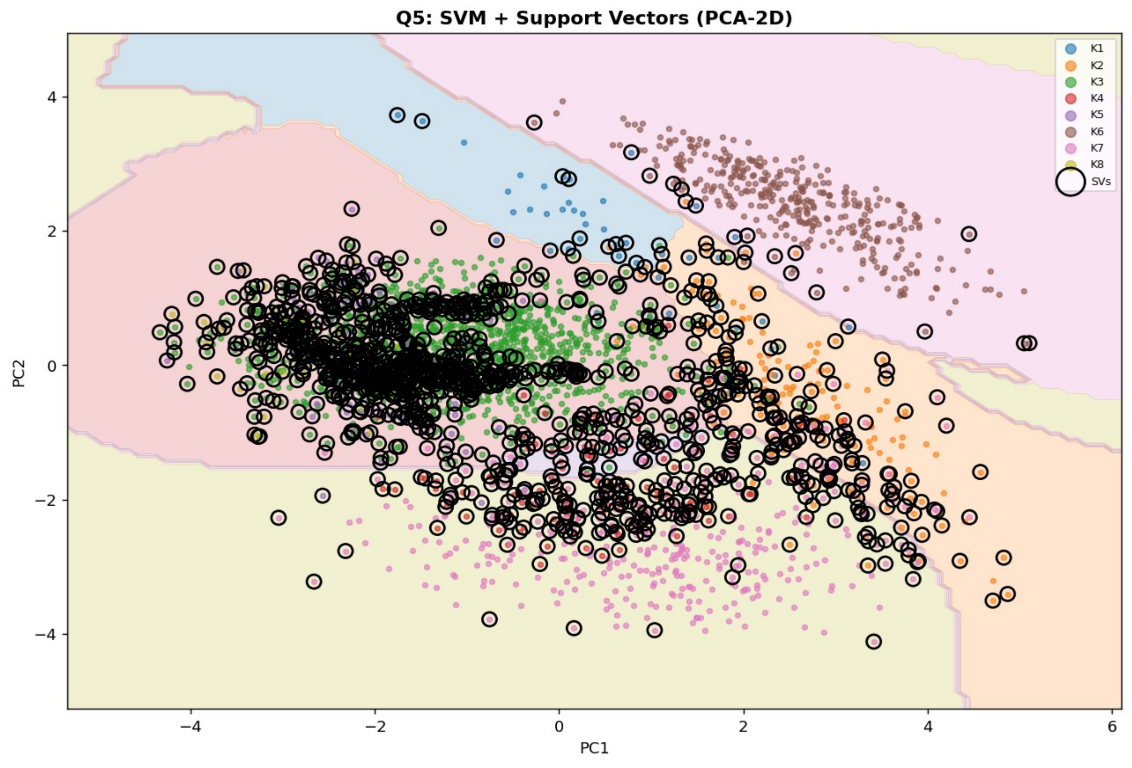
93.4%

Comparable
to Q4 Linear

SVM achieves ≈ same VAL as MLP but fits more tightly on TRAIN — the RBF kernel may be overfitting slightly.

Q5 — SVM Decision Regions & Support Vectors in PCA-2D

Q5



Non-linear boundaries

The RBF kernel creates curved, locally adaptive decision surfaces — impossible with a linear model.

Support vectors (○)

Hollow circles mark the SVs: the critical incidents lying closest to the decision boundary.

Boundary density

SVs concentrate most densely at K3/K6 and K3/K7 borders — the hardest classification zones.

Both the confusion matrix and VAL=93.4% confirm SVM is competitive with — but doesn't surpass — Logistic Regression.

Q6 — Multi-Layer Perceptron: Architecture & Training

Q6

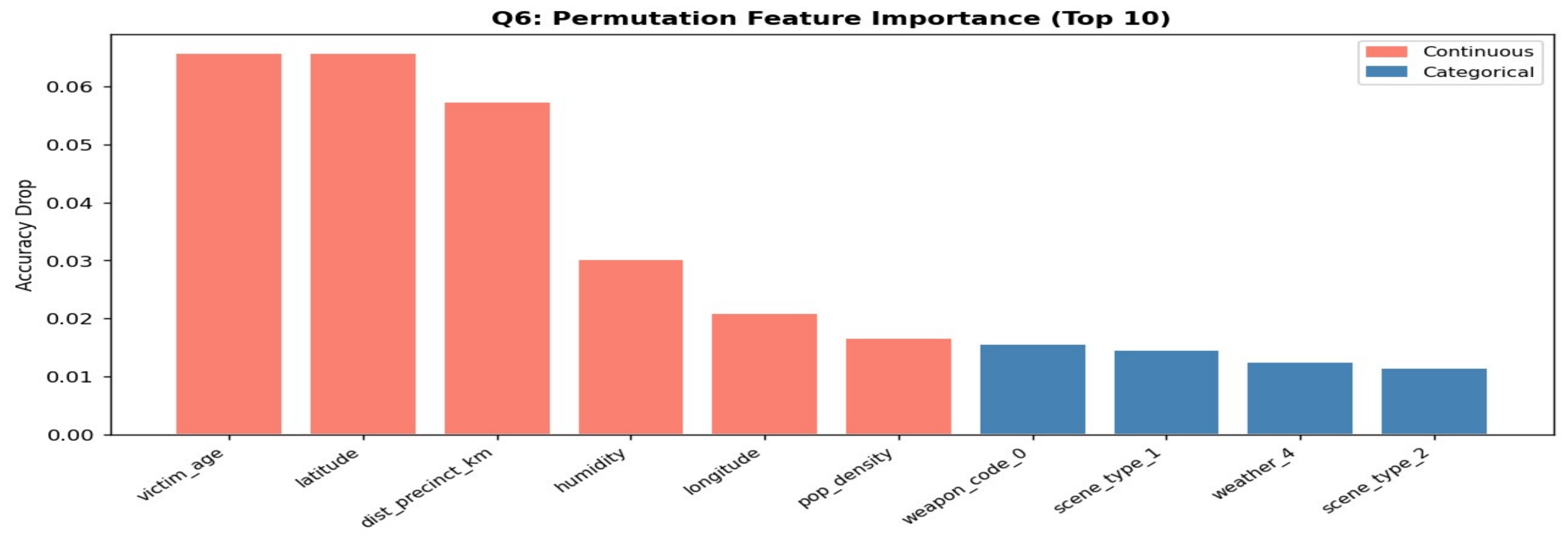
Input	Hidden 1	Hidden 2	Hidden 3	Output
25 units	128 units	64 units	32 units	8 units
	ReLU	ReLU	ReLU	Softmax

- Loss: Categorical Cross-Entropy
- Optimizer: Adam (lr = 10^{-3})
- Weight Decay: $\alpha = 10^{-3}$
- Early Stopping on 10% internal VAL

TRAIN	VAL
95.6%	93.4%

Q6 — Permutation Feature Importance: What Identifies a Killer?

Q6



#1 victim_age

Most powerful single feature — each killer targets a specific age profile

#2 latitude

Geographic territory — strong killer-specific spatial clustering

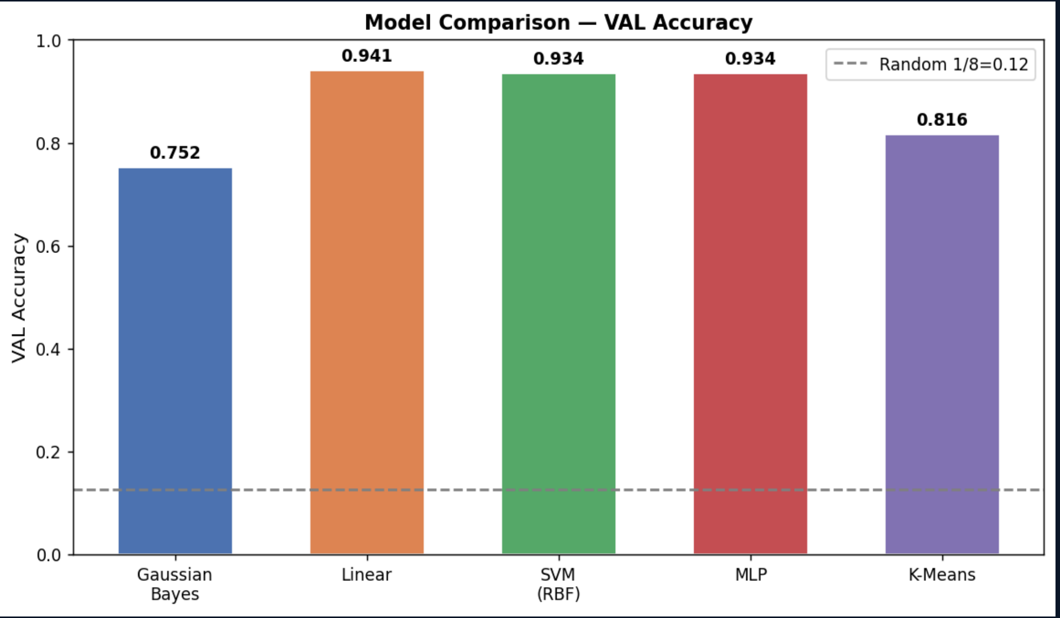
#3 dist_precinct_km

Operational risk tolerance — some killers avoid precincts, others don't

All top-5 features are continuous. Categorical features matter collectively but less individually.

Supervised Model Comparison — VAL Accuracy

Summary



Gaussian Bayes — **75.2%**

Continuous only. Gaussian assumption holds partially.

Linear (LR) — **94.1%**

Best model. Categorical features decisive.

SVM RBF — **93.4%**

Competitive. Mild overfit on TRAIN.

MLP 3-layer — **93.4%**

Same as SVM. Early stopping prevents overfit.

Linear decision boundaries suffice — the feature space, once encoded, is approximately linearly separable.

Q7 — Principal Component Analysis: Finding the Latent Structure

Q7

Goal: project the 25-dimensional feature space onto a lower-dimensional 'modus operandi' space that preserves maximum variance and reveals killer cluster structure.

PCA: eigen-decompose $\hat{\Sigma} = (1/N)\, \tilde{X}^T \tilde{X} = V \Lambda V^T \quad \rightarrow \quad z_i = V_m^T \tilde{x}_i \in \mathbb{R}^m$

Fitted on TRAIN only. Applied identically to VAL and TEST (no data leakage).

25

Input dimensions (standardised)

14

Components for 90% variance

75.6%

Variance in first 10 PCs

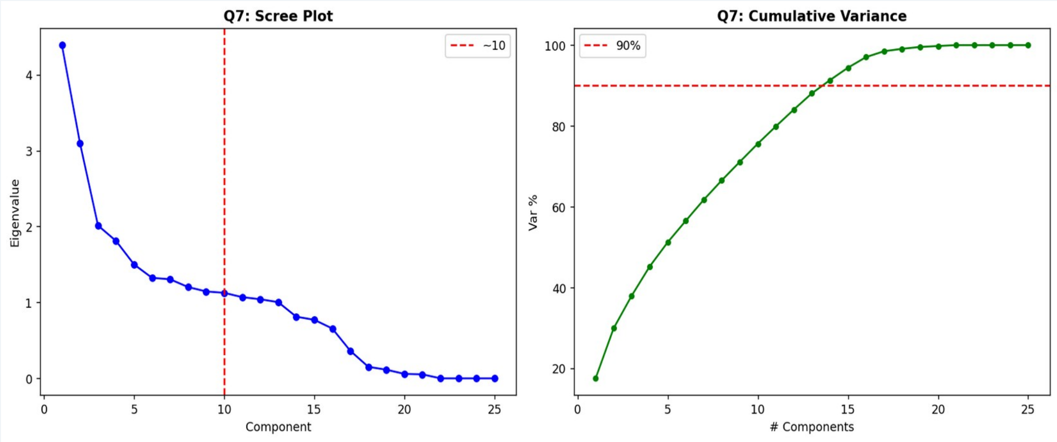
m=14

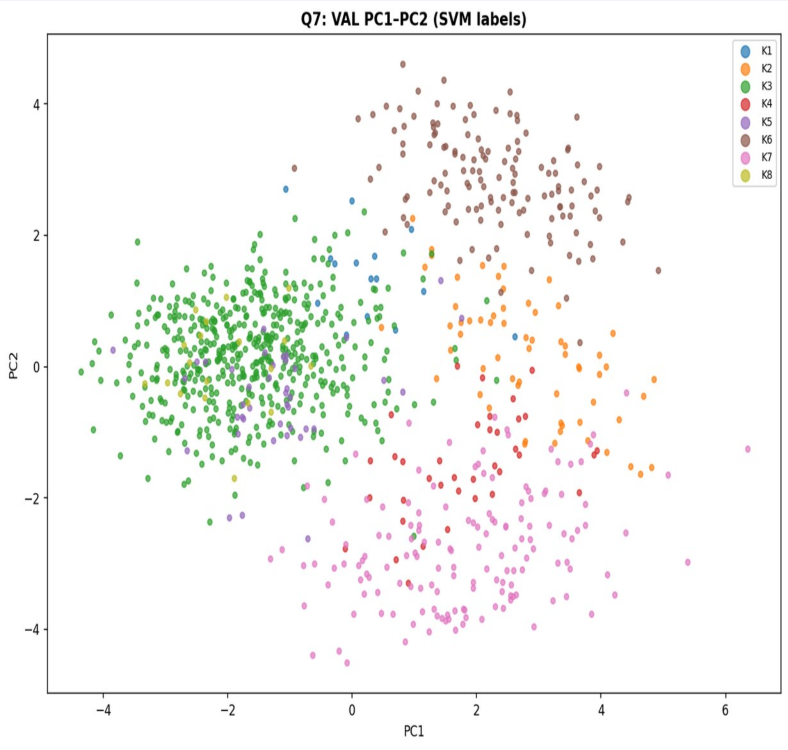
Selected latent dimension for Q8

The gradual elbow in the scree plot at ~10 components suggests no sharp dimensionality boundary — variance is spread across features.

Q7 — Scree Plot & VAL Killer Clusters in PCA-2D Space

Q7





Scree plot (left): elbow at ~10, 90% threshold at m=14 components selected.

- ~8 distinct clusters visible in just 2 PCs
- K3 occupies the largest central region
- Minority killers (K1, K8) form tight clusters

Scatter = VAL | Color = SVM predicted killer

Q8 — k-Means Clustering: Unsupervised Killer Attribution

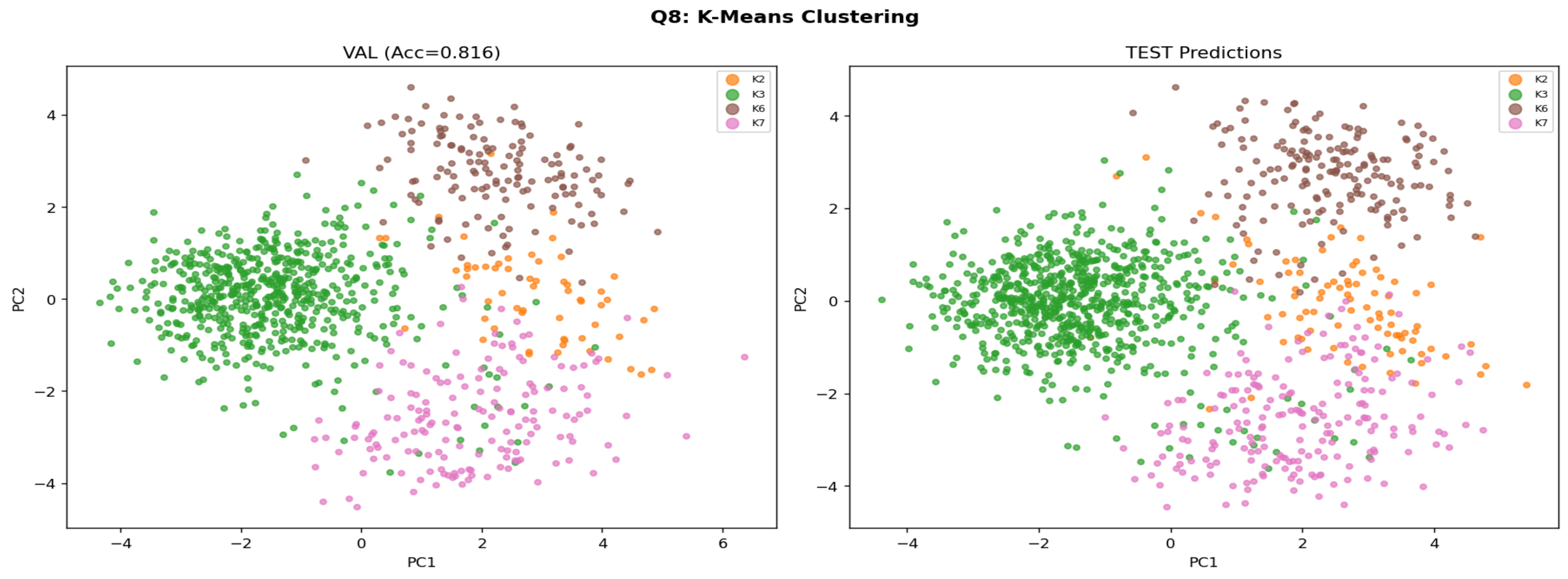
Q8

Step 1	Step 2	Step 3	Step 4
Project	**Cluster**	**Map**	**Predict**
Apply PCA-14 to all incidents (mean/directions from TRAIN only)	k-Means (k=8) on TRAIN PCA vectors k-means++ init, 20 restarts	Majority-vote: each cluster q gets the killer most frequent in it g(q) = argmax_k Σ 𝟙[c_i=q, K_i=k]	Assign VAL/TEST incident to nearest cluster, then look up g(q)

No killer labels are used at test time — only at the mapping step (TRAIN only). Purely unsupervised inference.

Q8 — K-Means: 81.6% VAL Accuracy Without Test Time Labels!

Q8



81.6%

VAL accuracy — unsupervised k-means beats the fully-supervised Bayes (75.2%)

Limitation: Multiple clusters mapped to K3 (the dominant class) — minority killers K1, K4, K5, K8 share a cluster or get absorbed.

Key Forensic Insights — What the Data Tells Us

Insights

📍 Killers have fixed spatial territories

Latitude and longitude are the #2 most important features. Each killer operates in a geographically coherent zone visible even in unsupervised clustering.

🎯 Victim age is the strongest single predictor

The top permutation-importance feature — each killer consistently targets victims in a specific age range (young adults vs. elderly), likely reflecting opportunistic targeting patterns.

🔫 Categorical features collectively decisive

The +18.9pp jump from Bayes → Linear Regression demonstrates that weapon, scene type, and weather conditions together define a killer's modus operandi.

🧩 Feature space is approximately linearly separable

The best model is the simplest (Logistic Regression). Non-linear models (SVM, MLP) offer no improvement, suggesting class boundaries are linear in the 25D feature space.

The 81.6% unsupervised k-means result confirms: killer identity is geometrically encoded in the feature space.

Conclusions & Deliverables

Best Model

Logistic Regression

VAL = 94.1%

Best Unsupervised

k-Means (PCA-14)

VAL = 81.6%

Top Feature

victim_age

Perm. imp. #1

Submission

submission.csv

4,800 predictions

The Piraeus Vice dataset is geometrically structured: killer identity is linearly separable in the 25-dimensional feature space.

Deliverables: solution_Q1_Q8.py · piraeus_vice_report.pdf · submission.csv · this slide deck