

Data Analysis Report

Generated: 2025-04-28 00:22:25

Dataset Information

shape: (733863, 30)
columns: ['msisdn', 'activation_date', 'subscriber_wilaya', 'subscriber_id_number', 'pdv_id', 'pdv_wilaya', 'status', 'customer_type', 'first_cdr_wilaya', 'subscriber_wilaya_std', 'pdv_wilaya_std', 'subscriber_wilaya_code', 'pdv_wilaya_code', 'activation_hour', 'activation_dayofweek', 'activation_month', 'activation_week', 'time_phase', 'same_wilaya', 'wilaya_code_diff', 'pdv_activation_count', 'days_since_pdv_first_activation', 'time_phase_code', 'subscriber_sim_count', 'pdv_total_activations', 'pdv_unique_customers', 'pdv_avg_sims_per_customer', 'pdv_multi_sim_pct', 'reconstruction_error', 'is_anomaly']
dtypes: {'msisdn': dtype('int64'), 'activation_date': dtype('O'), 'subscriber_wilaya': dtype('O'), 'subscriber_id_number': dtype('O'), 'pdv_id': dtype('O'), 'pdv_wilaya': dtype('O'), 'status': dtype('O'), 'customer_type': dtype('O'), 'first_cdr_wilaya': dtype('O'), 'subscriber_wilaya_std': dtype('O'), 'pdv_wilaya_std': dtype('O'), 'subscriber_wilaya_code': dtype('float64'), 'pdv_wilaya_code': dtype('int64'), 'activation_hour': dtype('int64'), 'activation_dayofweek': dtype('int64'), 'activation_month': dtype('int64'), 'activation_week': dtype('int64'), 'time_phase': dtype('O'), 'same_wilaya': dtype('int64'), 'wilaya_code_diff': dtype('float64'), 'pdv_activation_count': dtype('int64'), 'days_since_pdv_first_activation': dtype('int64'), 'time_phase_code': dtype('int64'), 'subscriber_sim_count': dtype('int64'), 'pdv_total_activations': dtype('int64'), 'pdv_unique_customers': dtype('int64'), 'pdv_avg_sims_per_customer': dtype('float64'), 'pdv_multi_sim_pct': dtype('float64'), 'reconstruction_error': dtype('float64'), 'is_anomaly': dtype('bool')}

missing_values: {'msisdn': 0, 'activation_date': 0, 'subscriber_wilaya': 0, 'subscriber_id_number': 0, 'pdv_id': 0, 'pdv_wilaya': 0, 'status': 0, 'customer_type': 0, 'first_cdr_wilaya': 0, 'subscriber_wilaya_std': 0, 'pdv_wilaya_std': 0, 'subscriber_wilaya_code': 14, 'pdv_wilaya_code': 0, 'activation_hour': 0, 'activation_dayofweek': 0, 'activation_month': 0, 'activation_week': 0, 'time_phase': 0, 'same_wilaya': 0, 'wilaya_code_diff': 14, 'pdv_activation_count': 0, 'days_since_pdv_first_activation': 0, 'time_phase_code': 0, 'subscriber_sim_count': 0, 'pdv_total_activations': 0, 'pdv_unique_customers': 0, 'pdv_avg_sims_per_customer': 0, 'pdv_multi_sim_pct': 0, 'reconstruction_error': 0, 'is_anomaly': 0}

memory_usage: 555.5298519134521
file_path: C:/Users/Zack/Desktop/PFE TESTING ON DATA/Processed CSV In Chunks/enhanced_data_with_anomalies_part1.csv

Statistical Summary

	mean	median	std	var	min	max	skew
	324843728.3687	324867641.0000	14373082.2221	206585492563066.7188	300010025.0000	349740621.0000	0.0076
	21.6994	19.0000	11.8962	141.5205	0.0000	99.0000	0.5660
	20.9668	19.0000	11.1782	124.9531	0.0000	58.0000	0.4459
	15.6471	16.0000	4.1652	17.3489	0.0000	23.0000	-0.5390
	3.0098	3.0000	2.0172	4.0690	0.0000	6.0000	-0.0012

	2.1959	2.0000	0.9192	0.8450	1.0000	4.0000	0.1830
	8.1705	8.0000	3.6485	13.3117	2.0000	15.0000	0.0496
	0.9063	1.0000	0.2914	0.0849	0.0000	1.0000	-2.7888
	1.8163	0.0000	7.0913	50.2868	0.0000	87.0000	4.5914
	1133.2545	440.0000	1796.7977	3228482.0140	1.0000	10846.0000	3.0629
	42.2077	41.0000	25.6627	658.5735	0.0000	89.0000	0.1104
	1.1932	1.0000	0.7136	0.5093	0.0000	3.0000	-0.2419
	4.2911	1.0000	61.9563	3838.5820	1.0000	1745.0000	27.5368
	1133.2545	440.0000	1796.7977	3228482.0140	1.0000	10846.0000	3.0629
	840.2472	354.0000	1236.8951	1529909.6037	1.0000	6726.0000	2.7637
	1.2997	1.1923	1.2882	1.6595	1.0000	29.6000	20.4851
	16.2311	14.7218	10.9413	119.7118	0.0000	100.0000	1.1010
	6.8032	0.0163	189.7962	36022.5897	0.0002	5744.2118	28.4081