KrispCall Revenue Attribution Analysis: Assignment Description

1 Objective

The KrispCall Revenue Attribution Analysis is a data analyst assessment designed to evaluate the ability to analyze customer data and attribute revenue to integration platforms using various attribution models. The goal is to provide actionable insights for optimizing platform partnerships, marketing strategies, customer success initiatives, and product development for KrispCall, a cloud-based communication platform integrated with 25 platforms (e.g., Salesforce, Monday, Zoho).

2 Business Context

KrispCall enables businesses to make calls and send SMS through web, mobile, and third-party integrations. Understanding which platforms drive revenue is critical for:

- Partnership Strategy: Prioritizing integration development.
- Marketing Investment: Allocating spend across platform ecosystems.
- Customer Success: Identifying high-value integration patterns.
- Product Development: Focusing on revenue-driving integrations.

3 Datasets

The analysis uses four datasets, detailed below:

3.1 Business Rules

- Only workspace owners generate revenue (payments/refunds).
- Any workspace user can connect integrations or make calls/SMS.
- Integration platform usage requires an active connection.
- Events must occur after first payment and before refunds.

4 Deliverables

The assignment requires the following outputs:

1. first_touch_attribution.csv: Revenue attributed to the first integration connected post-payment.

2. usage_based_attribution.csv: Revenue distributed by communication volume (call duration + SMS segments).
3. chosen_attribution.csv: Revenue attributed using a chosen model (e.g., Linear).
4. analysis_summary.md: 500-1000 word report explaining methodology, findings, and recommendations.
5. attribution_analysis.py: Complete Python code for the analysis.

5 Attribution Models

5.1 First-Touch Attribution

- Assigns 100% of net revenue to the first integration connected after the initial payment.
- Attributes to "No Integration" for customers without connections.

5.2 Usage-Based Attribution

- Distributes revenue based on communication volume (call duration in seconds + SMS segments normalized to 60 seconds per segment).
- Includes web and mobile platforms.

5.3 Chosen Model (Linear)

- Equally distributes revenue across all active integrations per customer.
- Justify selection and compare with other models in the summary.

6 Technical Requirements

- Data Validation: Verify user existence, temporal constraints, and platform availability.
- Attribution Calculations: Handle edge cases (no platform usage, zero revenue), ensure proper data types, round revenue to 2 decimal places.
- Code Quality: Use clear variable names, comprehensive comments, and error handling.
- Visualizations: Generate plots (e.g., histograms, count plots) and a comparison chart.

7 Results Comparison

The attribution models yield distinct insights into platform contributions to revenue, with a total net revenue of $4,775. The comparison table below summarizes revenue for key platforms across the models.

7.1 Comparison of Results

The First-Touch model attributes the majority of revenue ($2,335) to "No Integration", highlighting that many customers do not connect integrations after their initial payment. Among integrations, monday ($585) and pipedrive ($545) receive the highest attribution, indicating these are common initial connections that drive acquisition.

In contrast, the Usage-Based model shifts focus to actual communication activity, with web ($1,824.29) and mobile ($1,671.71) dominating due to their always-available nature and high usage volume (e.g., 1572 events, primarily calls and SMS). Integration platforms like zoho ($251.08), pipedrive ($197.33), and salesforce ($184.67) show significant contributions, suggesting these are heavily used for ongoing engagement.

The Linear model provides a balanced view, distributing revenue equally among all active integrations for each customer. "No Integration" still leads ($1,715), but platforms like monday ($360.83), pipedrive ($308.33), and asana ($273.33) receive more even attribution, rewarding broad adoption. This model reduces the dominance of non-integration platforms and highlights a wider range of integrations (e.g., 20 platforms receive revenue).

Overall, First-Touch emphasizes acquisition (initial integrations), Usage-Based reflects engagement (daily usage), and Linear promotes diversity (multiple integrations). Discrepancies, such as zoho's high Usage-Based but zero First-Touch attribution, suggest it is adopted later but used intensely.

7.2 Comparison of Plots

The visualizations provide supporting insights:

- Correlation Heatmap (correlation_heatmap.png): Displays weak correlations between payment/refund amounts and timestamps, indicating refunds are not strongly tied to payment size or timing, which supports the net revenue calculation approach.
- Total Users Histogram (total_users_histogram.png): Shows a right-skewed distribution (mean 2.08, median 1), confirming KrispCall's focus on small teams, aligning with high web/mobile usage in the Usage-Based model.
- Platform Count Plot (platform_count_plot.png): Illustrates the top 10 platforms, with web and mobile leading, followed by zoho and pipedrive. Labels are rotated 45 degrees for clarity, highlighting usage patterns that explain the Usage-Based model's results.
- Net Revenue Histogram (net_revenue_histogram.png): Reveals a right-skewed distribution ($0-$100 range), with outliers at higher values, explaining why "No Integration" dominates in First-Touch and Linear models (many low-revenue customers without integrations).
- Attribution Comparison Bar Plot (attribution_comparison.png): A horizontal bar plot comparing revenue across models, with green for First-Touch, blue for Usage-Based, and yellow for Linear. It visually emphasizes web/mobile dominance in Usage-Based and balanced distribution in Linear, aiding quick identification of key platforms.

These plots are generated using Matplotlib and Seaborn, with a Chart.js configuration for interactive rendering of the attribution comparison, allowing zooming and tooltips for detailed exploration.

8 Evaluation Criteria

Table 1: Data Dictionary

Dataset	Column	Description	Example
		final_workspace_data.csv	
	workspace_id	Unique workspace identifier	1, 2, 3
	workspace_name	Human-readable workspace name	SoloConsultant1
	total_users	Total users in workspace	1, 3, 8
	owner_email	Email of workspace owner	owner1@soloconsultant1.com
	member_emails	Comma-separated member emails	user1@company.com, user2@company.com
		final_payment_refund_data.csv	
	event	Event type	New Payment Made, Refund Granted
	time	Unix timestamp	1746072000
	$email	Customer email (payments)	owner1@company.com
	Amount	Payment amount (USD)	50.0, 150.0
	Refund Amount	Refund amount (USD)	50.0
		final_integration_data.csv	
	event	Event type	Integration Connected, Disconnected
	time	Unix timestamp	1746331200
	distinct_id	User email	user1@company.com
	Integration Platform	Platform name	salesforce, hubspot
		final_calls_sms_detailed_data.csv	
	event	Communication type	Outbound Calls, Outbound SMS
	time	Unix timestamp	1746331200
	Cost	Event cost (USD)	0.599433
	Duration	Call duration (seconds)	431.0 seconds
	Email	User email	user1@company.com
	Platform	Platform used	web, mobile, salesforce
	Segments	SMS segments	1, 2, 3

Table 2: Revenue Attribution by Platform and Model (USD)

Platform	First-Touch	Usage-Based	Linear
No Integration	2335.00	0.00	1715.00
monday	585.00	174.67	360.83
pipedrive	545.00	197.33	308.33
asana	0.00	173.33	273.33
activecampaign	290.00	156.00	235.00
freshsales	175.00	178.67	210.00
salesforce	160.00	184.67	195.00
zoho	0.00	251.08	177.50
web	0.00	1824.29	0.00
mobile	0.00	1671.71	0.00
front	150.00	0.00	165.83
engagebay	160.00	110.67	125.00

Table 3: Evaluation Criteria

Category	Description	Weight
Technical Implementation	Correct model implementation, data handling, validation, code clarity	40%
Business Understanding	Accurate interpretation of requirements, edge case handling, revenue accuracy	30%
Analysis Quality	Clear methodology, insightful comparisons, actionable recommendations	20%
Communication	Professional writing, logical structure, effective visualizations	10%