

PART 1: Welcome to the World of Data Hackathons

Welcome to the world of Data Analytics hackathons! It is a different beast from software dev hackathons. In software dev, you build a *product* (an app, a bot, a tool). In data analytics, you build a *story* backed by evidence.

Here is the breakdown of what this world is about and specifically what **10Alytics** looks for.

1. The Core Objective: "From Data to Decision"

In a dev hackathon, if your code works and the UI is slick, you win. In analytics, "working code" is just the bare minimum.

The goal is to take a raw, messy dataset and extract **actionable intelligence**.

- **Dev Hackathon:** "Look, I built a drone delivery app."
- **Data Hackathon:** "I analyzed 10 years of logistics data and found that routing drones through *these* specific corridors saves 15% fuel. Here is the dashboard to monitor it, and here is the projected ROI."

2. What They Are Actually Looking For (The "Winning" Rubric)

Based on 10Alytics' past trends and general data hackathon standards, you are judged on these four pillars:

A. The Storytelling (Most Important)

This is where most devs fail. They show chart after chart. **Do not do this.**

- **Bad:** "Here is a bar chart of sales by region. Here is a line chart of revenue."
- **Good:** "We noticed a 20% drop in revenue in Q3. Drilling down, we found it was specific to the North region. The data suggests this correlates with the new competitor entering that market. Therefore, we recommend..."
- **Tip:** Treat your presentation like a pitch deck to a CEO, not a code review.

B. The "So What?" (Actionable Insights)

10Alytics judges love **recommendations**.

- If your analysis shows that malaria cases are rising in a specific region, don't just stop there.
- **The Win:** "We recommend allocating 40% of the budget to mosquito nets in *Region X* specifically during *Month Y*, which data shows is the peak infection window."

C. Visual Excellence (UI/UX of Data)

- **Tools:** They often use Power BI, Tableau, or Python/R (Streamlit/Dash).
- **Expectation:** Your dashboards should be intuitive. A user should look at it and understand the trend in 5 seconds.
- **Clutter is fatal:** Avoid "chart junk" (too many colors, 3D pie charts, confusing labels).

D. Technical Rigor (The Engine Room)

- **Data Cleaning:** Real-world data is dirty. Show (briefly) that you handled missing values, duplicates, and outliers correctly.
- **Feature Engineering:** Did you create new useful variables from the existing data? (e.g., converting "Date of Birth" to "Age Group").

3. Specifics About 10Alytics Hackathons

From their previous editions, here is their specific "flavor":

- **Theme Heavy:** They lean heavily towards **Social Good** and **African Context**.
 - *Past Topics:* Health Crises (Malaria, Infant Mortality), Education in Africa, UN Sustainable Development Goals (SDGs).
- **The "Executive Summary" Matters:** They often require a one-page summary or a short video pitch.
- **Tools:** They are tool-agnostic but heavily favor **Power BI** and **Excel** for the final presentation layer, often supported by Python/SQL for the heavy lifting.

PART 2: The 2024 Edition Day 1 Analysis (The Survivor's Guide)

Since you are a software dev hackathon winner, you already know how to work under pressure. However, you need to shift your mindset from **"Building a Product"** to **"Selling a Solution based on Evidence."**

1. The Core Objective

Theme: "Tackling Child and Infant Mortality in Africa."

Goal: You aren't just visualizing data. You are acting as a consultant to the United Nations or African Governments. You must analyze the data to propose **actionable, data-driven strategies** to reduce preventable deaths in children under five.

2. The Winning Formula (Rubric)

According to the organizers (Chukwuemeka and the Judges), here is exactly how you are scored. This is different from code quality scoring:

1. **Understanding of the Case Study:** Did you actually grasp the problem, or did you just start making charts? You need to interpret the scenario correctly.
2. **Storytelling & Eloquence:** This is the #1 differentiator. Do not just say "Sales went up." Say "Because X happened, Y resulted, therefore we should do Z."
3. **Quality of Recommendations (The "So What?"):**
 - **Bad:** "Malaria is high in Nigeria."
 - **Winning:** "Allocating resources to mosquito nets in *Region X* during *Month Y* will reduce mortality by Z%."
 - **Key Term:** Move from *Descriptive Analytics* (what happened) to **Prescriptive Analytics** (what should we do).
4. **Visual Excellence:** Are your dashboards cluttered? Can a CEO understand them in 5 seconds?
5. **Technical Ability:** Your data cleaning, modeling, and usage of tools (Python, PowerBI, Excel, etc.).
6. **Creativity ("Break the Box"):** The CEO explicitly said, "Break the box." Don't just do the bare minimum. Bring in external context if allowed/relevant to support your argument.

3. The Dataset "Schema"

You were provided with **8 specific datasets** (likely CSV/Excel sheets). You will probably need to join these by "Country Code" or "Year" to find correlations.

1. **Health Protection Coverage:** % of population covered by health insurance.
2. **Global Vaccination Coverage:** Vax rates for various diseases.
3. **Births Attended by Skilled Staff:** % of births assisted by pros.
4. **Maternal Deaths by Region:** Estimated maternal deaths.
5. **Child Mortality by Income Level:** Under-5 mortality rates across income levels.
6. **Infant Deaths:** Annual number of infant deaths.
7. **Youth Mortality Rates:** Mortality for individuals under 15.
8. **Causes of Death in Children Under 5:** The breakdown of major causes.

Dev Tip: You are looking for **correlations**.

- *Example Hypothesis:* Does higher "Health Protection Coverage" correlate with lower "Infant Deaths"? Which specific vaccination has the highest impact on survival rate?

4. Logistics & Rules

- **Time Limit:** 10 Hours (12:00 PM – 10:00 PM WAT).
- **Tools:** Tool Agnostic. Use whatever you want (Python, R, Excel, Power BI, Tableau, Canva).
- **Submission:**
 - **One single entry.** Even if you have a dashboard and a slide deck, zip them or put the link to the dashboard *inside* the PDF/PPT.
 - **Format:** PPT, PDF, PowerBI file, or Excel.
 - **Deadline:** Hard stop at 10:00 PM WAT.

5. Strategic Advice from the Transcript (The "Secret Sauce")

- **Target Audience:** Design your presentation for **Policymakers and Stakeholders**, not for other data analysts. Keep it accessible.
- **The Flow:**
 - **Slide 1:** Executive Summary (The Problem).
 - **Slide 2-3:** The Insights (The Trends/Visuals).
 - **Slide 4:** The "Why" (Correlations/Drivers).
 - **Slide 5: The Solution (Recommendations).** *Spend the most energy here.*
- **Don't Over-Engineer:** A complex Python script that produces a confusing graph will lose to a simple Excel chart that clearly shows a life-saving insight.
- **Limitation Clause:** It is impressive if you include a slide on the "Limitations of the Data" (e.g., "Data for 2023 was missing, so we projected based on 2022").

6. How to Pivot as a Dev

- **Your Advantage:** You can likely clean data via Python/Pandas faster than the Excel guys. Use that speed to spend more time on the **Business Logic**.
 - **Your Weakness:** Devs tend to focus on the "Build." In this hackathon, the "Build" (the dashboard) is just a prop for the "Pitch." Focus on the Pitch.
-

PART 3: The 2024 Finale – Part 1 (Judges & Presenters)

This first part of the finale (Day 2) is essentially a masterclass in **what to do** and **what NOT to do**. Since you are a developer, you will appreciate this: **The code/dashboard is only 40% of the grade. The "Pitch" is the other 60%.**

1. The Judges & Their "Hot Buttons"

Knowing your audience is key. Here is what the specific judges looked for in this segment:

- **Olutoyin (Head of Data):** The **"So What?"** Guy.
 - *He hates:* Just showing numbers (e.g., "GDP is high").
 - *He loves:* Context. "GDP is high, *but* it doesn't correlate to health coverage because the wealth isn't distributed." He wants **Actionable Intelligence**.
- **Sunkanmi (The Storyteller):** The **"Pacing"** Guy.
 - *Feedback:* "Understand the **Power of the Pause**."
 - *Tip:* Don't rush through your charts. Show a chart, pause, let them absorb the insight, then move to the recommendation.
- **Sherifat (Data Governance):** The **"Data Quality"** Hawk.
 - *Hot Button:* **Missing Data**. She explicitly asked, "How did you handle the data gaps?"
 - *The Win:* Admitting you used the "Median" to fill gaps or explicitly stating the data was insufficient is better than pretending the data was perfect.
- **Favor (Data Scientist):** Looked for **Authenticity**. She wanted to see *your* logic, not just generic charts.

2. Analysis of the First 3 Presenters (The Good, The Bad, & The Buggy)

Presenter 1: Richard (The "ML" Guy)

- **The Approach:** He went technical. He built a **Random Forest Regression Model** to predict future mortality rates.
- **The Dev Mistake:** He focused on the *tool* (Machine Learning) rather than the *result*. The model had low predictive power, and he had to admit it during Q&A.
- **The Lesson:** Don't over-engineer. A simple bar chart that is 100% accurate beats a complex AI model that is 70% accurate. **Don't use a bazooka to kill a mosquito.**

Presenter 2: Oluwabamishe (The "Storyteller" - Strong Contender)

- **The Approach:** He used a "Drill Down" narrative.
 - Slide 1: Global View.
 - Slide 2: Africa View.
 - Slide 3: West Africa (The Problem Area).
 - Slide 4: Niger (The Specific Country).
- **Why Judges Loved It:** He didn't just show data; he led them down a funnel.
- **The "Dev" Win:** He used **Parameters** in Power BI (interactive buttons) to let judges switch views. This is the "UI/UX" of analytics.
- **Key Insight:** He proved that "Skilled Birth Attendants" (nurses/midwives) had a higher correlation to saving lives than just "GDP."

Presenter 3: Chinonso (The "Emotional Hook" - Scope Creep Error)

- **The Approach:** She started with an emotional hook ("Imagine losing a child...").
- **The Mistake:** The prompt was **"Africa."** She immediately narrowed down to **"Nigeria"** because she is Nigerian.
- **The Feedback:** The judges scolded her (politely). "You narrowed down to Nigeria too fast." You cannot solve an African problem by only looking at one country.
- **The Lesson: Read the Spec.** If the hackathon asks for a solution for *Region X*, do not build a solution for *Sub-region Y* just because it's easier or you have more data.

3. The "Dev-to-Analyst" Translation Guide (Actionable Steps)

A. Handling "Missing Data" (The Null Pointer Exception of Analytics)

In software, if a variable is null, the app crashes. In analytics, if a row is empty, you have to make a decision.

- **What to do:** If the prompt gives you dirty data (they will), do not delete the rows. Fill them with the **Median** or **Average** of that year, and **tell the judges you did it**.
- *Quote from Video:* "I used the median because the mean would skew the analysis." (This is a high-IQ answer).

B. The "Actionable Recommendation" Framework

When you present your final slide, use this formula which worked for the presenters:

1. **Insight:** "We see high mortality in rural areas."
2. **Driver:** "Data shows this is because of low vaccination rates, not lack of money."
3. **Action:** "Therefore, we recommend **Mobile Vaccination Clinics** rather than building new hospitals."

C. The "Power of Pause"

As a dev, you are used to demoing features fast ("Here's the login, here's the cart, here's the checkout").

- **Stop that.**
 - Display the dashboard.
 - Say: "If you look at the top left..."
 - **Wait 3 seconds.**
 - "...you will see a downward trend."
-

PART 4: Deep Dive – The "Hidden" Insights from Part 1

Upon re-analyzing the transcript with a fine-tooth comb—specifically looking for the *technical* and *strategic* nuances that might trip up a developer—there are **four critical "hidden" insights**. These are the specific traps you need to avoid.

1. The "Black Box" Trap (The Machine Learning Warning)

What Happened: The first presenter (Richard) used **K-Means Clustering** (an unsupervised learning algorithm) to group countries, and then a **Random Forest** (supervised) for prediction.

The Trap: One of the judges (Favor, the Data Scientist) asked a lethal question: "*How were you able to validate the quality of your clustering since you didn't use the target variable?*"

The Insight for You:

- **Don't use fancy algorithms just to impress.** If you use a "Black Box" model (like Neural Networks or Random Forest) in a hackathon, you **must** be able to explain *why* it works and *how* you validated it.
- **The Fix:** If you can't explain the math behind your validation in 30 seconds, stick to simple Regression or Correlation analysis. A simple model you understand is better than a complex one you fumble.

2. The "Transparency of Transformation" (The ETL Question)

What Happened: The second presenter (Oluwabamishe) mentioned he had to merge datasets that didn't quite fit (different codes/entities).

The Trap: Judge Sherifat asked specifically: *"I'd like to understand the transparency of how the data has been transformed... what gaps you filled by yourself and what impact it might have."*

The Insight for You:

- Judges don't just trust your charts. They want to know if you "cooked" the data.
- **The Fix:** Have a specific slide (or a dedicated 30 seconds) titled **"Data Processing Methodology."**
 - Say this: "The dataset had 40% missing values for [Year X]. I imputed these using the [Median/Linear Interpolation] method because [Reason]. This ensures the trend remains visible without skewing the average."
- **Dev translation:** This is your "Unit Test" defense. Prove your data logic is sound.

3. The "Lowest Hanging Fruit" Strategy

What Happened: Judge Sunkanmi praised one presenter for finding the "Lowest Hanging Fruit."

The Insight for You:

- You cannot solve *all* of Africa's mortality problems in one presentation.
- **The Winning Move:** Identify the **one** variable that is easiest to fix but has the highest impact.
 - Example: "We found that while building hospitals is expensive (High Effort), increasing DTP3 vaccination coverage by 10% reduces mortality by 15% (Low Effort, High Impact). Therefore, our primary recommendation is a **Vaccination Drive**, not infrastructure."
- **This demonstrates "Business Acumen,"** which judges value over raw coding skill.

4. The "Bias" Check (The Nigeria vs. Africa Issue)

What Happened: Presenter 3 (Chinonso) saw data that showed **Niger** and **Sierra Leone** had worse stats than **Nigeria**. However, she focused her presentation on **Nigeria**.

The Trap: Judge Sunkanmi called this out. *"There were other two countries ahead of us... but you hammered so much on Nigeria."*

The Insight for You:

- **Avoid "Home Country Bias."** Do not focus on your country just because you live there or understand it. Follow the data.
- **The Fix:** If the data says Sierra Leone is the problem, build your solution for Sierra Leone. If you *must* pivot to your home country, give a data-backed reason (e.g., "While Sierra Leone has higher rates, Nigeria has a higher *population volume*, meaning more total lives can be saved here").

Summary of Added "Cheat Codes":

1. **Validation:** If you ML, explain your validation metrics (R-Squared, RMSE, Silhouette Score).

2. **ETL Transparency:** Admit where the data was bad and how you fixed it.
 3. **Prioritization:** Recommend the "Low Effort / High Impact" solution first.
 4. **Objectivity:** Follow the data, not your bias.
-

PART 5: The 2024 Finale – Part 2 (The Winner vs. The Losers)

This second part is even more revealing than the first because it shows the **extreme contrast** between a winning presentation and a losing one. It offers a perfect "Don't Do This" list.

1. The Star Performance: Daniel (The "Consultant")

Daniel is the benchmark you need to aim for. He didn't just present data; he pitched a business case.

- **The "Hook":** He started with a high-level economic view (Income levels) and drilled down to the specific health metric.
- **The "One-Thing" Question:** When asked to pick **only one** solution out of four, he didn't hesitate. He picked "Economic Growth" because his data showed it was the root cause of the other issues.
 - *Lesson:* Judges will force you to prioritize. Don't say "everything matters." Pick one driver and defend it with data.
- **Visuals:** Clean, simple charts. No walls of text.
- **The Result:** The judges had almost no criticism, only praise.

2. The Cautionary Tales (How to Lose Points Fast)

Presenter 2: Atiejoye (The "Tech Failure")

- **The Disaster:** He spent 3-4 precious minutes struggling to share his screen and turn on his camera.
- **The Presentation:** Once he started, he simply read the text on his slides.
- **The "Dev" Trap:** As a dev, you know "it works on my machine" is not a valid excuse.
 - *The Fix:* Test your Zoom/Teams setup 30 minutes before. Have a PDF version of your slides ready to send to the host if your screen share fails.
- **The Verdict:** He lost the audience before he even showed a chart.

Presenter 3: Mustafa (The "Lecturer")

- **The Mistake: Death by Powerpoint.** His slides were 80% text and 20% small charts. He spent his time reading the paragraphs to the judges.
- **The Consequence:** He ran out of time. The moderator had to cut him off *before* he could finish his recommendations (the most important part).
- **The Statistical Error:** He presented a correlation of **-0.3** (which is mathematically weak) but tried to frame it as a significant finding.
 - *Judge's Catch:* He admitted the relationship was "not very strong" because the R-Squared value was low (56% or 0.56).
 - *Lesson:* If your R-Squared is low, admit the model is weak or look for other variables. Don't pretend a weak correlation is a strong insight.

3. Critical Feedback from the Judges (The Rubric in Action)

Judge Sherifat (The Hawk) on "Data Pedigree"

She caught Mustafa on a critical error: **Data Consistency**.

- **The Issue:** The dataset provided had gaps. Mustafa introduced "Per Capita Income" data but couldn't clearly explain if it was in the original dataset or if he brought it in from an external source (World Bank).
- **The Rule:** If you bring in external data (which is allowed and encouraged), **cite your source clearly**.
 - *Say this:* "The hackathon dataset didn't have income data, so I enriched it with World Bank Open Data to find this correlation."

Judge Sunkanmi on "Visual Hierarchy"

- **The Issue:** Mustafa used a heat map that was cluttered and hard to read.
- **The Lesson:** If a Judge has to squint to read your axis labels, you have failed.
 - *Dev Tip:* Think of this like UI design. If the user (Judge) can't find the button (Insight) in 2 seconds, the UX is bad.

4. Your "Hackathon Cheat Codes" for the Final Preparation

Based on this segment, here are 3 things you must do to beat 90% of the participants:

1. **The "10-Word Rule" for Slides:**

- Never have a paragraph on a slide. Use bullet points.
- *Better yet:* Put the text in the "Speaker Notes" and only put the **Chart** and a **Big Bold Headline** on the screen.
- *Why:* If they are reading your slide, they aren't listening to you.

2. Master the "Prioritization Matrix":

- Daniel won because he linked **Income -> Health**.
- Don't just list 5 problems. Find the **Root Cause**.
- *Formula:* "While factors A, B, and C contribute to mortality, our analysis shows that **Factor A** is the upstream driver. Fixing A makes fixing B and C easier."

3. The "Timebox" Strategy:

- Mustafa got cut off.
- If you have 10 minutes, practice for **8 minutes**.
- Structure:
 - Min 0-2: Problem & Context.
 - Min 2-6: Data Analysis & Visuals (The "Meat").
 - Min 6-8: Recommendations & Conclusion (The "Money").
 - Min 8-10: Buffer/Q&A prep.

PART 6: The Official Winning Blueprint (Deep Dive & Strategy)

This is the "money" round. Since you have won ML hackathons before, you know that winning is about **pattern recognition**. You look at what the judges reward and what they punish, and you optimize for that.

I have mapped the **Official Criteria** directly to the **evidence** from the video so you know exactly what they mean by those terms.

Criterion 1: UNDERSTANDABILITY

Definition: Understanding the problem and framing the statement.

The Hidden Insight from Part 2:

Most devs frame the problem as "The data is messy, and I need to clean it." **This is wrong.**

- **The Loser's Move (Mustafa):** He spent precious time showing slides titled "Data Preprocessing," "Data Integration," and "Feature Engineering." He explained *how* he cleaned the data. **The judges were bored.**
- **The Winner's Move (Daniel):** He framed the problem as an **Economic Issue**. He didn't talk about null values; he talked about *poverty cycles* causing death.
- **The "Dev" Pivot:** Do the cleaning (you are good at this), but **hide the kitchen**. Do not show the dirty dishes. Show the meal. Your problem statement should be about *Human Lives*, not *CSV files*.

Criterion 2: INNOVATIVENESS

Definition: Quality of insights extracted from the dataset.

The Hidden Insight from Part 2:

Innovation in analytics often means **"Data Enrichment."**

- **The Masterstroke (Daniel):** The provided dataset likely didn't have "Income Classification" (Low/Middle/High income) for every country. Daniel **went to the World Bank**, got external data, merged it, and found his winning insight: *Income is the #1 driver of mortality*.
- **The Judge's Reaction:** Judge Olu specifically asked, "*Was per capita income given, or did you infer it?*" When Daniel cited his external source, he gained massive credibility.
- **Action Item:** Don't just use the csv they give you. Find one relevant external dataset (e.g., GDP, Literacy Rate, Number of Doctors per 1000 people) and merge it. This counts as "Innovativeness."

Criterion 3: IMPACTFULNESS

Definition: Completeness of insights and solutions/recommendations.

The Hidden Insight from Part 2:

Impact is measured by **Prioritization** (The "Boil the Ocean" trap).

- **The Trap:** Ebube gave a laundry list of recommendations: "Drones, education, sanitation, better hospitals..." It felt like a wish list.
- **The Winning Move:** Judge Sunkanmi backed Daniel into a corner: "*If you had to pick ONE factor, what would it be?*"
 - Daniel didn't waffle. He picked **"Economy."** He argued that if you fix the economy, the hospitals and sanitation fix themselves.
- **Action Item:** In your recommendations, use a **"Priority Matrix."**
 - "We have 5 recommendations, but the **'Big Domino'** is [X]. If we push [X], the others fall into place."

Criterion 4: APPLICABILITY

Definition: Enthusiasm and presentation of findings.

The Hidden Insight from Part 2:

This is where the "ML/Dev" mindset often fails.

- **The "Death by PowerPoint" (Mustafa):** He had slides with 3 charts and 4 paragraphs of text. He ran out of time reading the text. The judges cut him off before his recommendations. **Automatic Loss.**
 - **The "Visual Anchor" (Daniel):** He used **large, single charts** per slide. He spoke to the chart, he didn't read *from* the slide.
 - **The "Tech Check" (Ebube):** She lost 3 minutes fumbling with screen share and video. It killed her momentum and "Enthusiasm" score.
 - **Action Item:**
 - **One Chart Per Slide.**
 - **Font Size 24+ only.**
 - **No Code.** (Unless it's a snippet proving a complex algorithm, but even then, hide it).
-

Specific Advice for You (The "ML Winner" Strategy)

Since you have won ML hackathons, you have a secret weapon, but it's a double-edged sword.

1. The "Black Box" Risk

Ebube ran a **Random Forest Classifier** (78% accuracy).

- **The Problem:** In an *Analytics* hackathon, "Accuracy" doesn't matter as much as "Explainability." If you can't explain *why* the model predicts death, the policy-makers (judges) won't trust it.
- **Your Strategy:** If you run an ML model, focus on **Feature Importance**.
 - *Don't say:* "My model is 85% accurate."
 - *Do say:* "My Random Forest model identified that '**Distance to Water Source**' is the #1 predictor of mortality, outweighing even 'Vaccination Status'." -> *This is an insight.*

2. The "Correlation vs. Causation" Trap

Mustafa showed a correlation matrix with a score of **-0.3** (weak correlation) and tried to sell it as a strong finding. The judges caught him immediately.

- **Your Strategy:** You know math. If the correlation is weak, **admit it.**

- *Winning Line*: "We looked for a correlation between X and Y, but the data shows it's actually weak. This contradicts popular belief and suggests we should look at Z instead." -> *Negative results are also insights.*

3. The "Confidence" Factor

You have experience winning. Use that stage presence.

- **Judge Sunkanmi's Feedback:** He explicitly praised "The Power of the Pause."
- **Technique:** When you reveal your main insight, **stop talking for 3 seconds**. Let them look at the chart. Then deliver the recommendation.

Summary Checklist for Your Deck:

1. **Slide 1:** Title + "The Big scary Number" (e.g., "1.2 Million Children Die Every Year").
2. **Slide 2:** The Root Cause (Data Viz proving specific driver, e.g., Income or Sanitation).
3. **Slide 3:** The External Validation (Your "Innovative" external dataset).
4. **Slide 4:** The "Big Domino" Recommendation (The one thing that fixes everything).
5. **Slide 5:** The Impact (Projected lives saved if they listen to you).