Certify Intel Final Excel Deliverable

Complete User Guide

Document Version: 1.0

Last Updated: January 18, 2026

Prepared For: Certify Health Team

PART 1: WHAT IS THIS EXCEL FILE?

Simple Explanation

This Excel file is your **Competitive Intelligence Command Center**. Think of it like a smart filing cabinet that:

1. **Stores information** about all your competitors (85+ companies)
2. **Automatically finds** new competitors you might not know about
3. **Alerts you** when something important changes
4. **Creates reports** you can share with your team

Instead of manually searching websites and taking notes, this spreadsheet does most of the work for you.

The File You Will Use

File Name: `Certify_Intel_ExcelOnly_v3_1_dashboard_ops_READY_CLEAN.xlsx`

Where to Find It: `C:\Users\conno\Downloads\Certify_Health_Intelv1\`

File Size: About 1.6 MB (megabytes)

PART 2: THE 25 TABS EXPLAINED

When you open the Excel file, you will see 25 different tabs (also called "sheets") at the bottom of the screen. Here is what each one does:

CATEGORY 1: SETUP TABS (6 Tabs)

These tabs control how the system works. You only need to change these occasionally.

Tab 1: Brand_Kit

What it is: A reference page showing Certify Health's brand colors and fonts.

What you do with it: Look at it when you want to make sure your reports match the company style. You don't need to change anything here.

In simple terms: It's like a style guide poster hanging on the wall.

Tab 2: Config

What it is: The settings page for the entire system.

What you do with it: This is where you enter your API keys (special passwords that let the system connect to outside services).

Important cells:

- **Cell B4:** Enter your Bing Search API key here
- **Cell B5:** Enter your SerpAPI key here (optional)
- **Cell B6:** Maximum number of competitors to track (default is 50)
- **Cell B9:** How many days before data is considered "stale" (default is 30)

In simple terms: Like the control panel on your car dashboard.

Tab 3: Lists

What it is: Dropdown menus that appear in other parts of the spreadsheet.

What you do with it: You don't usually edit this directly. It automatically pulls information from other tabs.

In simple terms: Like the dictionary the system uses to make dropdown menus.

Tab 4: Sources_Allowlist

What it is: A security list of approved websites the system can collect information from.

What you do with it: If you want the system to read a new website, add it here first.

Columns you'll see:
- **source_key:** The category of the source (like "Company_Site")
- **enabled:** Set to TRUE to allow, FALSE to block
- **domain:** The website address (like "phreesia.com")

Why this matters: This prevents the system from accidentally going to bad websites. If a website isn't on this list, the system won't touch it.

In simple terms: Like a VIP list for a club - only approved sites get in.

Tab 5: Segment_Rules

What it is: Definitions for how competitors are grouped into market segments.

What you do with it: Define what makes a "Small Business" vs "Enterprise" competitor, or what qualifies as "Patient Intake" vs "Revenue Cycle."

In simple terms: Like sorting your contacts into different folders.

Tab 6: Field_Dictionary

What it is: The master list of every type of information the system tracks.

What you do with it: Look here to understand what each data point means.

Example row:
- pricing_acv_mid_market | Pricing | number | USD/year | 0.8 | TRUE

This means: "Annual pricing for mid-market customers, measured in dollars per year, needs 80% confidence to be trusted, and must come from a primary source."

In simple terms: Like a glossary at the back of a textbook.

Tab 7: Feature_Taxonomy

What it is: Standard names for product features.

What you do with it: This ensures that "online check-in," "digital intake," and "web registration" are all recognized as the same feature.

In simple terms: Like a translator that makes sure everyone uses the same words.

CATEGORY 2: DATA ENTRY TABS (3 Tabs)

These are where the actual competitor information lives.

Tab 8: Input (THE MAIN DATA TAB)

What it is: This is the heart of the system - where all competitor information is stored.

This tab contains 3 tables:

Table 1: tblEntities (Your Competitor List)

Every competitor gets one row here:

Column	What It Means	Example
entity_id	Unique ID number	E001
canonical_name	Company name	Phreesia
domain	Website	phreesia.com
vertical	Industry category	Patient Intake
region	Where they operate	US
segment_id	Market segment	SEG_MM (Mid-Market)
discovered_via	How we found them	BING_SEARCH or DEMO
discovered_at	When we found them	2026-01-12
status	Are they active?	ACTIVE
hq_city	Headquarters city	New York

Color coding:

- **Green cells** = Data imported automatically
- **Blue cells** = Data you entered manually

Table 2: tblEvidence (Where We Got the Info)

Every piece of information has a source recorded here:

Column	What It Means
evidence_id	Unique ID for this evidence
entity_id	Which competitor this is about
source_key	Type of source (Company website, news article, etc.)
tier	TIER1 (direct from company) or TIER2 (third-party)
url	The web address where we found this
fetched_at	When we grabbed this information
snippet	The actual text we captured

Why this matters: You can always trace back where any piece of data came from.

Table 3: tblClaimCandidates (Possible Data Points)

When the system finds information, it goes here first for review:

Column	What It Means
candidate_id	Unique ID
field_key	What type of data (like "pricing_acv")
value_text	The text value
value_num	The number value (if applicable)
confidence	How sure we are this is correct (0% to 100%)
status_auto	Did the system approve it?
status_final	Final decision: PROMOTED (accepted) or REVIEW_REQUIRED

In simple terms: This is the "inbox" where new information waits to be verified.

Tab 9: Claims_Current

What it is: Only verified, trusted data points appear here.

What you do with it: This is the "clean" data you can trust for reports.

In simple terms: Information that passed quality control.

Tab 10: Claim_Versions

What it is: A complete history of every change ever made.

What you do with it: Look here to see what data looked like in the past.

In simple terms: Like a time machine for your data. You can see every version of every data point.

CATEGORY 3: CHARTS AND VISUALIZATIONS (5 Tabs)

These tabs show your data in visual form.

Tab 11: Metrics

What it is: Calculated numbers like averages, totals, and percentages.

What you do with it: View this to see high-level statistics about your competitors.

In simple terms: Your data summary page.

Tab 12: Tech_Dashboard

What it is: A technical overview with 2 charts.

Charts included:
1. **Pie Chart** - Shows distribution (like how many competitors are in each category)
2. **Bar Chart** - Shows comparisons (like which competitors have the most funding)

Tab 13: Dashboard (MAIN DASHBOARD)

What it is: Your executive summary view with 2 charts.

Charts included:

1. **Bar Chart** - Competitor comparison bars
2. **Scatter Chart** - Two-dimensional analysis (each dot is a competitor)

Buttons on this page:

• **"Run Pipeline"** - Click to refresh all data
• **"Run Self Tests"** - Click to check if everything is working

Tab 14: Feature_Matrix

What it is: A grid comparing which features each competitor has.

How to read it:

• Rows = Competitors (one row per company)
• Columns = Features (one column per feature)
• Cells = YES (has it), NO (doesn't have it), or UNKNOWN

In simple terms: A checklist showing who has what.

Tab 15: Competitor_Detail

What it is: Deep-dive profiles for individual competitors.

What you do with it: Look here for detailed information about one specific company.

CATEGORY 4: WORKFLOW TABS (6 Tabs)

These tabs track tasks and changes.

Tab 16: Review_Queue

What it is: Your to-do list of data that needs human review.

Columns you'll see:

Column	What It Means

task_id	Unique ID for this task
created_at	When the task was created
task_type	What kind of review is needed
priority	HIGH, MEDIUM, or LOW
entity_id	Which competitor
field_key	What data point needs review
confidence	How confident the system is
status	OPEN, IN_PROGRESS, RESOLVED, or REJECTED

How to use it:

1. Sort by priority (HIGH first)
2. Look at the data in question
3. Change status to RESOLVED (if correct) or REJECTED (if wrong)

Tab 17: Events

What it is: A log of every change the system detected.

What you'll see:

Column	What It Means
event_id	Unique ID
event_type	What happened (like "EVIDENCE_CHANGED")
severity	How important (medium, high)
url	Which webpage changed
detected_at_utc	When we noticed the change
summary	Brief description
status	OPEN (not addressed) or CLOSED (addressed)

In simple terms: A news feed of competitor changes.

Tab 18: Alerts

What it is: Rules for what should trigger notifications.

How to use it:

- Set `enabled` to TRUE for alerts you want
- Set `enabled` to FALSE for alerts you don't want

Example alert: "Tell me whenever any competitor's website content changes"

Tab 19: Evidence_Snapshots

What it is: Saved copies of webpages over time.

Why this matters: You can compare what a competitor's website said last week vs. today.

Tab 20: Logs

What it is: General system activity log.

Tab 21: Run_Log

What it is: History of every time you ran the data refresh pipeline.

Columns:

Column	What It Means
run_id	Unique ID
started_at_utc	When it started
mode	manual or scheduled
status	running, completed, or failed
notes	Any errors or messages

CATEGORY 5: SYSTEM TABS (4 Tabs)

These help you set up and maintain the system.

Tab 22: Scraper

What it is: Settings for automated web scraping.

Tab 23: Scripts

What it is: Copy-paste code for the automation features.

What you'll find here:
- VBA code (for macros)
- Power Query code (for connecting to the web API)
- Step-by-step setup instructions

Tab 24: Self_Tests

What it is: Automated checks that verify everything is working correctly.

How to read it:
- PASS = Everything is fine
- FAIL = Something needs attention

Tab 25: Macro_Install

What it is: Instructions for enabling the automation features.

In simple terms: A user manual for turning on the advanced features.

PART 3: THE CHARTS EXPLAINED

The Excel file contains **4 charts** that visualize your data:

Chart 1: Pie Chart (Tech_Dashboard tab)

What it shows: How your competitors are divided into categories.

How to read it: Each colored slice represents a different category. Bigger slices = more competitors in that group.

Example: If the "Patient Intake" slice is biggest, that means most of your tracked competitors are in patient intake.

Chart 2: Bar Chart (Tech_Dashboard tab)

What it shows: Comparison of competitors on a specific metric.

How to read it: Each bar represents one competitor. Taller bars = higher values.

Example: If the chart shows "Employee Count," taller bars mean bigger companies.

Chart 3: Bar Chart (Dashboard tab)

What it shows: Quick comparison of key metrics across competitors.

How to read it: Look for which competitors have the tallest bars - those are the leaders in that category.

Chart 4: Scatter Chart (Dashboard tab)

What it shows: Two-dimensional comparison of competitors.

How to read it:
- Each dot represents one competitor
- Horizontal position (left to right) = one metric (like funding)
- Vertical position (bottom to top) = another metric (like employees)
- Dots in the upper-right corner = high on both metrics

Example: A dot in the upper-right means a company with lots of funding AND lots of employees.

PART 4: THE 17 DATA TABLES EXPLAINED

The file contains 17 named tables. Here's what each one stores:

Table Name	What It Stores
tblSourcesAllowlist	List of approved websites
tblSegmentRules	Market segment definitions
tblFieldDictionary	Data point specifications
tblFeatureTaxonomy	Standard feature names
tblEntities	Competitor master list
tblEvidence	Source citations
tblClaimCandidates	Unverified data points
tblClaimsCurrent	Verified data points
tblMetrics	Calculated statistics
tblReviewQueue	Human review tasks
tblLogs	System activity log
tblClaimVersions	Data change history
tblEvidenceSnapshots	Saved webpage copies
tblEvents	Change notifications
tblAlerts	Alert rules
tblRunLog	Pipeline execution history
tblFeatureMatrix	Feature comparison grid

PART 5: THE AUTOMATION FEATURES (VBA MACROS)

The Excel file can be automated using macros. Here's what each one does:

How to Enable Macros

1. Save the file as `.xlsm` (Excel Macro-Enabled Workbook)
2. Press `Alt + F11` to open the code editor
3. Go to **Insert → Module**
4. Paste the code from the `/vba/` folder
5. Close the editor and save

The 12 Automation Modules

Module 1: CertifyIntel_Core

What it does: Basic building blocks used by all other modules.

Key functions:

- Generates unique IDs for new records
- Gets the current time
- Finds tables by name

You don't need to change this.

Module 2: CertifyIntel_HTTP

What it does: Connects to websites and downloads data.

You don't need to change this.

Module 3: CertifyIntel_Allowlist

What it does: Checks if a website is approved before visiting it.

Security feature: If a website isn't on the allowlist, the system won't go there.

Module 4: CertifyIntel_FetchExtract

What it does: Visits competitor websites and extracts information.

How it works:

1. Goes to each competitor's website
2. Saves a copy of the page content
3. Looks for patterns (like prices)
4. Creates data candidates for review

Module 5: CertifyIntel_Discovery_Bing

What it does: Searches the internet for new competitors automatically.

How to use it:

1. Enter your Bing API key in the Config tab
2. Enter your search query (or use the default)
3. Run the pipeline

Example search: "Certify Health competitors patient intake healthcare"

Module 6: CertifyIntel_Review

What it does: Creates tasks in the Review Queue when data needs human verification.

Module 7: CertifyIntel_Versions

What it does: Saves approved data to the permanent history.

Module 8: CertifyIntel_EventsAlerts

What it does: Detects when competitor information changes and creates notifications.

How it works:

1. Compares today's saved webpages to yesterday's
2. If different, creates an event
3. Checks if any alerts should fire

Module 9: CertifyIntel_RunPipeline

What it does: Runs all the automation steps in order.

The steps:

1. Search for new competitors
2. Visit competitor websites
3. Extract data
4. Save approved data
5. Detect changes
6. Create alerts
7. Refresh everything

How to use it: Click the "Run Pipeline" button on the Dashboard tab.

Module 10: CertifyIntel_UI

What it does: Creates the buttons on the Dashboard tab.

Module 11: CertifyIntel_SelfTests

What it does: Runs the validation tests to check if everything is working.

How to use it: Click the "Run Self Tests" button on the Dashboard tab.

Module 12: JsonLite

What it does: Reads data in JSON format from web APIs.

You don't need to change this.

PART 6: CONNECTING TO THE WEB API (Power Query)

You can connect the Excel file to the backend server to get live data.

How to Set Up Power Query

1. Start the backend server:

• Open a command prompt

• Type: `cd C:\Users\conno\Downloads\Certify_Health_Intelv1\backend`

• Type: `python -m uvicorn main:app --port 8000`

2. In Excel:

• Go to **Data → Get Data → From Web**

• Enter: `http://localhost:8000/api/export/json`

• Click OK

3. Transform the data:

• Click on "competitors" to expand

• Click "To Table"

• Expand all columns

• Click "Close & Load"

4. Set up auto-refresh (optional):

• Right-click the table

• Select Properties

• Check "Refresh every X minutes"

PART 7: API KEYS NEEDED

To use all the features, you need these API keys:

Required Keys (Without these, core features won't work)

Key Name	What It Does	Where to Get It
OPENAI_API_KEY	Powers AI data extraction	https://platform.openai.com/api-keys
BING_API_KEY	Finds new competitors automatically	https://portal.azure.com

Recommended Keys (Makes the system much better)

Key Name	What It Does	Where to Get It
CRUNCHBASE_API_KEY	Funding and investor data	https://www.crunchbase.com/api
NEWSAPI_KEY	News article monitoring	https://newsapi.org

Optional Keys (Advanced features)

Key Name	What It Does	Where to Get It
SIMILARWEB_API_KEY	Website traffic data	https://www.similarweb.com/corp/developer/
LINKEDIN_API_KEY	Employee and hiring data	https://developer.linkedin.com
SERPAPI_KEY	Alternative search (more reliable)	https://serpapi.com

Email Notification Settings

Setting	What It Does	Default Value
SMTP_HOST	Email server address	smtp.gmail.com
SMTP_PORT	Email server port	587
SMTP_USER	Your email address	(you enter this)
SMTP_PASSWORD	Your email app password	(you enter this)

PART 8: HOW TO DO COMMON TASKS

Task 1: Add a New Competitor Manually

1. Go to the **Input** tab
2. Find the **tblEntities** table
3. Scroll to the bottom and add a new row
4. Fill in at least: `entity_id`, `canonical_name`, `domain`, `status` = ACTIVE

Task 2: Refresh All Data

1. Go to the **Dashboard** tab
2. Click the **"Run Pipeline"** button
3. Wait for it to complete (check Run_Log for status)

Task 3: Review Flagged Data

1. Go to the **Review_Queue** tab
2. Sort by priority (HIGH first)
3. For each item:
• Look at the data and source
• If correct: Change status to RESOLVED
• If wrong: Change status to REJECTED

Task 4: Check for Competitor Changes

1. Go to the **Events** tab
2. Look for OPEN status events
3. Review the changes
4. Mark as CLOSED when addressed

Task 5: Export Data to Share

Option A: Download Excel
• Go to `http://localhost:8000/api/export/excel` in your browser
• Save the downloaded file

Option B: Get PDF Report
• Go to `http://localhost:8000/api/reports/comparison` in your browser
• Save the PDF

PART 9: TROUBLESHOOTING

Problem: "Run Pipeline" Button Doesn't Work

Solutions:

1. Make sure the file is saved as `.xlsm` (macro-enabled)
2. Enable macros when prompted
3. Make sure VBA modules are imported

Problem: No Data Showing in Charts

Solutions:

1. Check if the Input tab has data in tblEntities
2. Run the pipeline to refresh data
3. Check if filters are hiding data

Problem: Power Query Says "Unable to Connect"

Solutions:

1. Make sure the backend server is running
2. Check the URL is correct: `http://localhost:8000`
3. Check if your firewall is blocking port 8000

Problem: Emails Not Sending

Solutions:

1. Check SMTP settings in the `.env` file
2. For Gmail, make sure you're using an App Password (not your regular password)
3. Test with: `python test_email.py`

PART 10: GLOSSARY

Term	Simple Definition
API Key	A password that lets programs talk to each other
Allowlist	A list of approved items (like approved websites)
Claim	A piece of data about a competitor
Confidence	How sure we are that data is correct (0-100%)
Entity	A competitor company
Evidence	Where we got a piece of information from
Macro	A program inside Excel that automates tasks
Pipeline	A series of automated steps that run in order
Power Query	Excel's tool for connecting to online data
Scraping	Automatically reading information from websites
Snapshot	A saved copy of a webpage at a specific time
Table	An organized grid of data in Excel
Tier	How trustworthy a source is (Tier 1 = most trusted)
VBA	The programming language used for Excel macros

DOCUMENT END

Total Tabs: 25

Total Data Tables: 17

Total Charts: 4

Total VBA Modules: 12

For questions or support, contact your system administrator.

Document prepared by Antigravity Agent

January 18, 2026