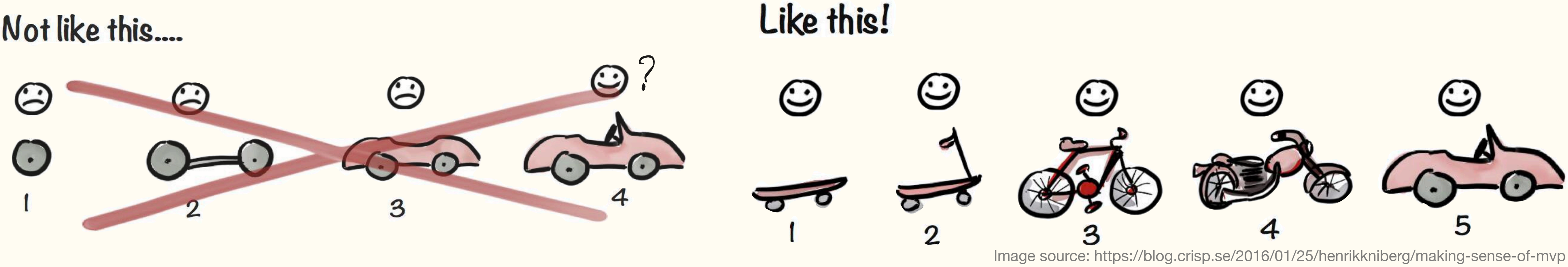


# The MVP Approach to Building a Data-Science Product


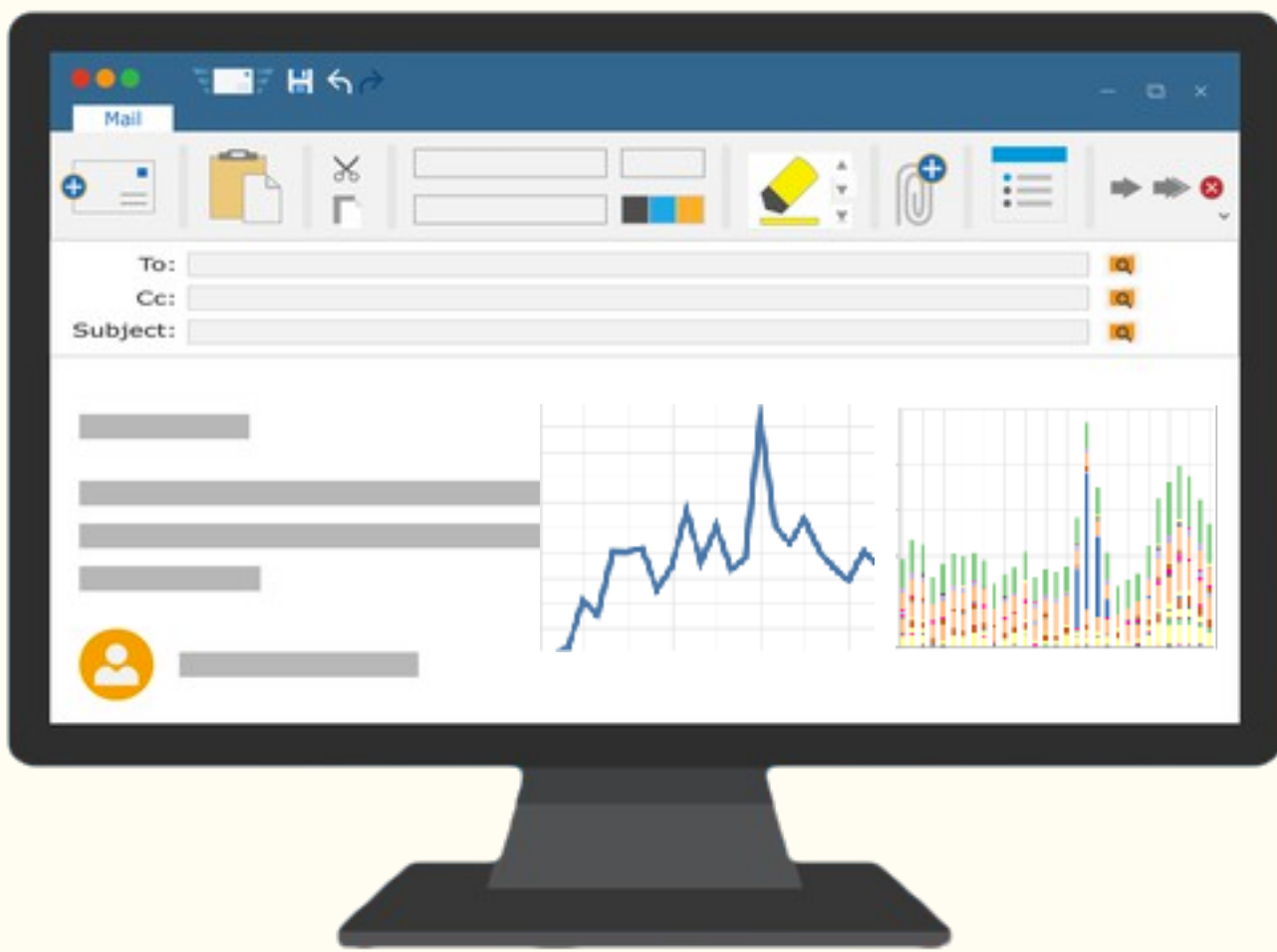
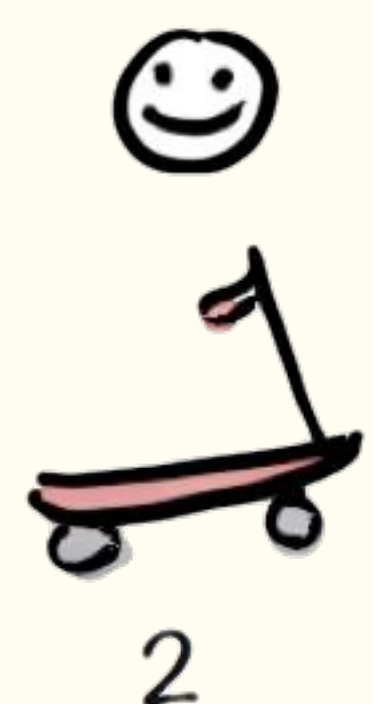


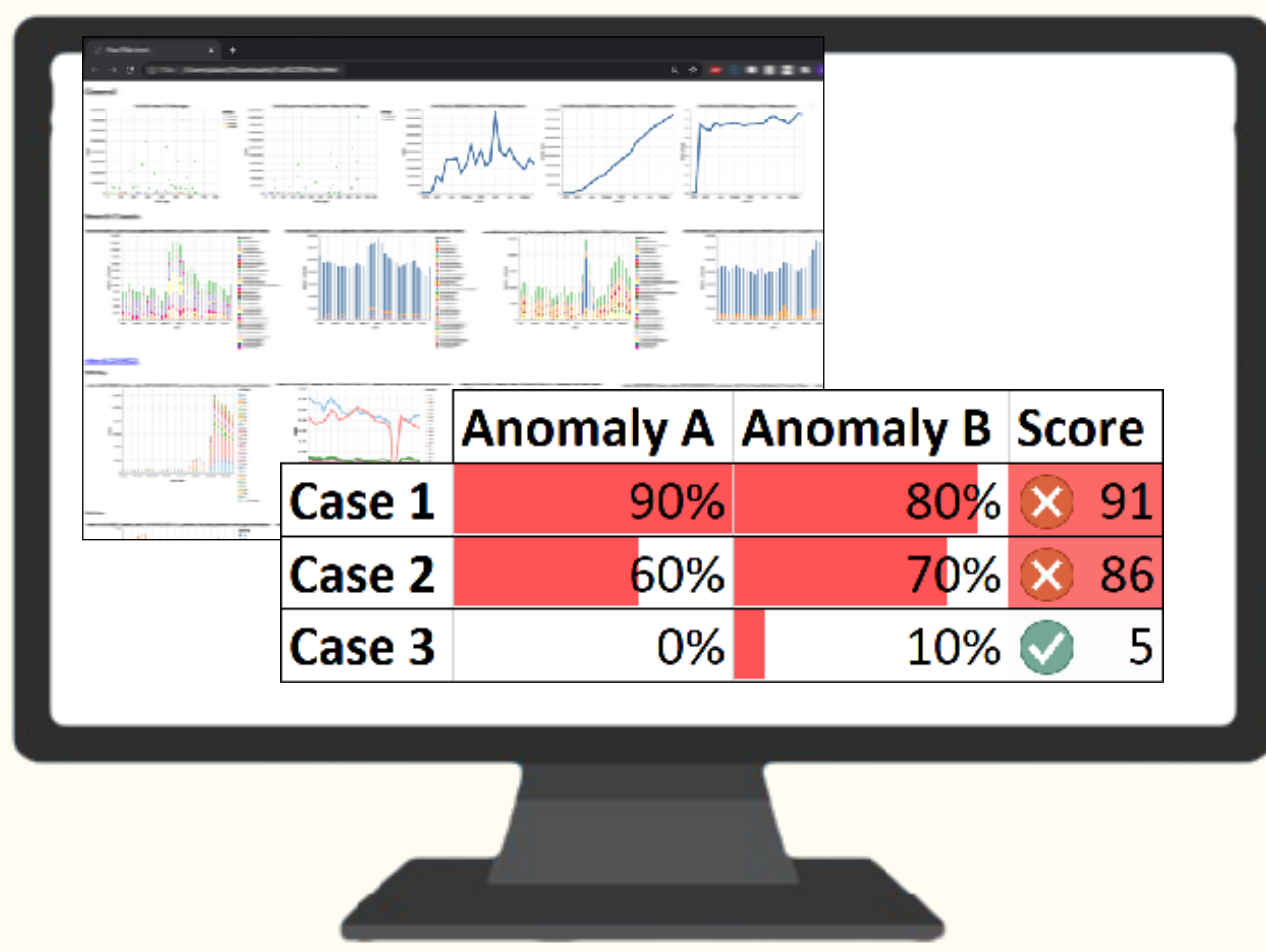

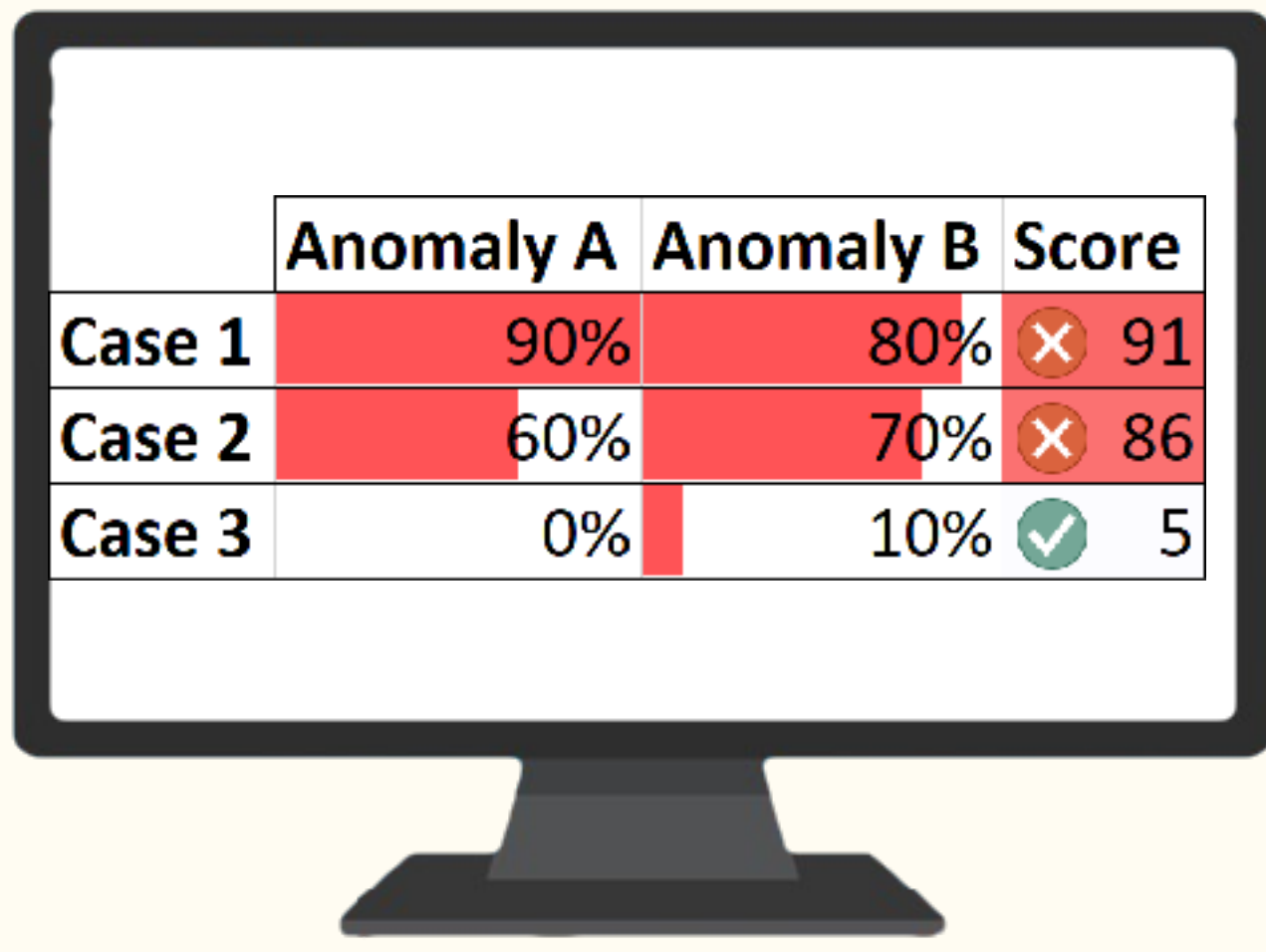

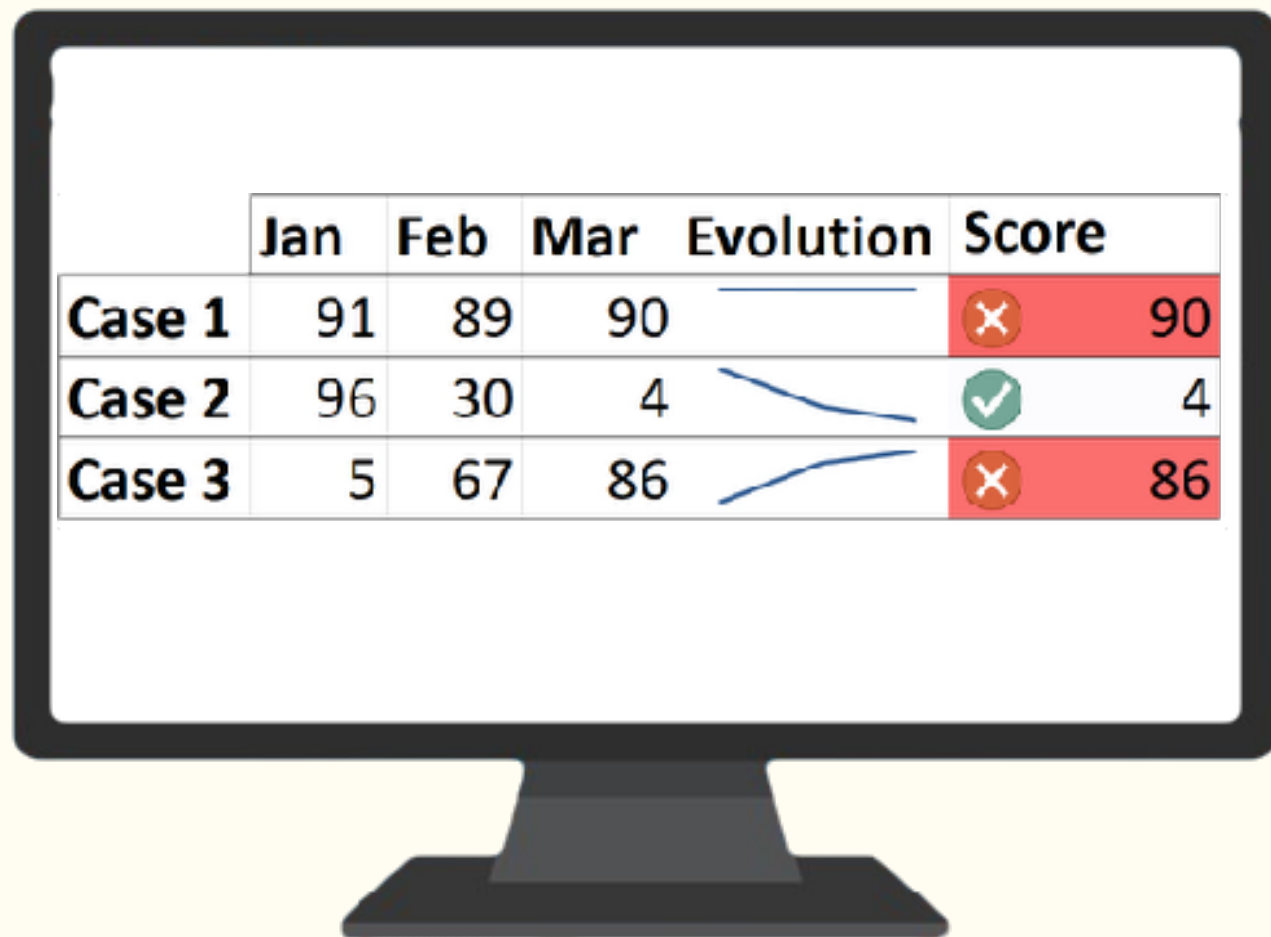
Author: [linkedin.com/in/alainsaas/](https://www.linkedin.com/in/alainsaas/)  
[twitter.com/alainsaas](https://twitter.com/alainsaas)



Starting with a Minimal Viable Product, followed by incremental iterations, is a common approach in software product development.

How does it translate to data-science products?

Let's take a new fraud detection system as example:

			Analysis	Decision Making	Explainability																								
 1		First analyses conducted on individual cases are sent to Account Management by email.  <b>Goal:</b> Engage stakeholders	Case-by-case	Manual	Detailed reports																								
 2		Systematic dashboards are developed to study and report cases.  <b>Goal:</b> Test the impact of actions taken on users, to confirm ROI	Case-by-case	Systematic	Detailed reports																								
 3	 <table><tr><th></th><th>Anomaly A</th><th>Anomaly B</th><th>Score</th></tr><tr><td>Case 1</td><td>90%</td><td>80%</td><td>91</td></tr><tr><td>Case 2</td><td>60%</td><td>70%</td><td>86</td></tr><tr><td>Case 3</td><td>0%</td><td>10%</td><td>5</td></tr></table>		Anomaly A	Anomaly B	Score	Case 1	90%	80%	91	Case 2	60%	70%	86	Case 3	0%	10%	5	Algorithms and machine learning are trained to detect patterns using cases studied previously manually, and then applied to new cases.  Manual reviews are still conducted on edge cases to tweak the algorithms.  <b>Goal:</b> Scale up	Cluster-by-cluster + edge cases	Study of patterns at scale	Detailed reports for edge cases								
	Anomaly A	Anomaly B	Score																										
Case 1	90%	80%	91																										
Case 2	60%	70%	86																										
Case 3	0%	10%	5																										
 4	 <table><tr><th></th><th>Anomaly A</th><th>Anomaly B</th><th>Score</th></tr><tr><td>Case 1</td><td>90%</td><td>80%</td><td>91</td></tr><tr><td>Case 2</td><td>60%</td><td>70%</td><td>86</td></tr><tr><td>Case 3</td><td>0%</td><td>10%</td><td>5</td></tr></table>		Anomaly A	Anomaly B	Score	Case 1	90%	80%	91	Case 2	60%	70%	86	Case 3	0%	10%	5	Algorithms are run massively to detect all existing cases, leveraging historical data.  <b>Goal:</b> Finish solving the existing problem	Automated on existing cases	Systematic	Auditable								
	Anomaly A	Anomaly B	Score																										
Case 1	90%	80%	91																										
Case 2	60%	70%	86																										
Case 3	0%	10%	5																										
 5	 <table><tr><th></th><th>Jan</th><th>Feb</th><th>Mar</th><th>Evolution</th><th>Score</th></tr><tr><td>Case 1</td><td>91</td><td>89</td><td>90</td><td></td><td>90</td></tr><tr><td>Case 2</td><td>96</td><td>30</td><td>4</td><td></td><td>4</td></tr><tr><td>Case 3</td><td>5</td><td>67</td><td>86</td><td></td><td>86</td></tr></table>		Jan	Feb	Mar	Evolution	Score	Case 1	91	89	90		90	Case 2	96	30	4		4	Case 3	5	67	86		86	Additional developments are done to allow an automated monitoring of new cases and their evolution over time.  <b>Goal:</b> Prevent the problem from reappearing	Automated on new cases	Systematic	Auditable
	Jan	Feb	Mar	Evolution	Score																								
Case 1	91	89	90		90																								
Case 2	96	30	4		4																								
Case 3	5	67	86		86																								