

Quickプレ診断書_サマリ

システム毎/カラム毎に、問題データ(難易度高)/問題データ(難易度低)/有効データの割合を掲載する。各データの詳細については、グラフ単位で次頁以降に診断結果として掲載する。

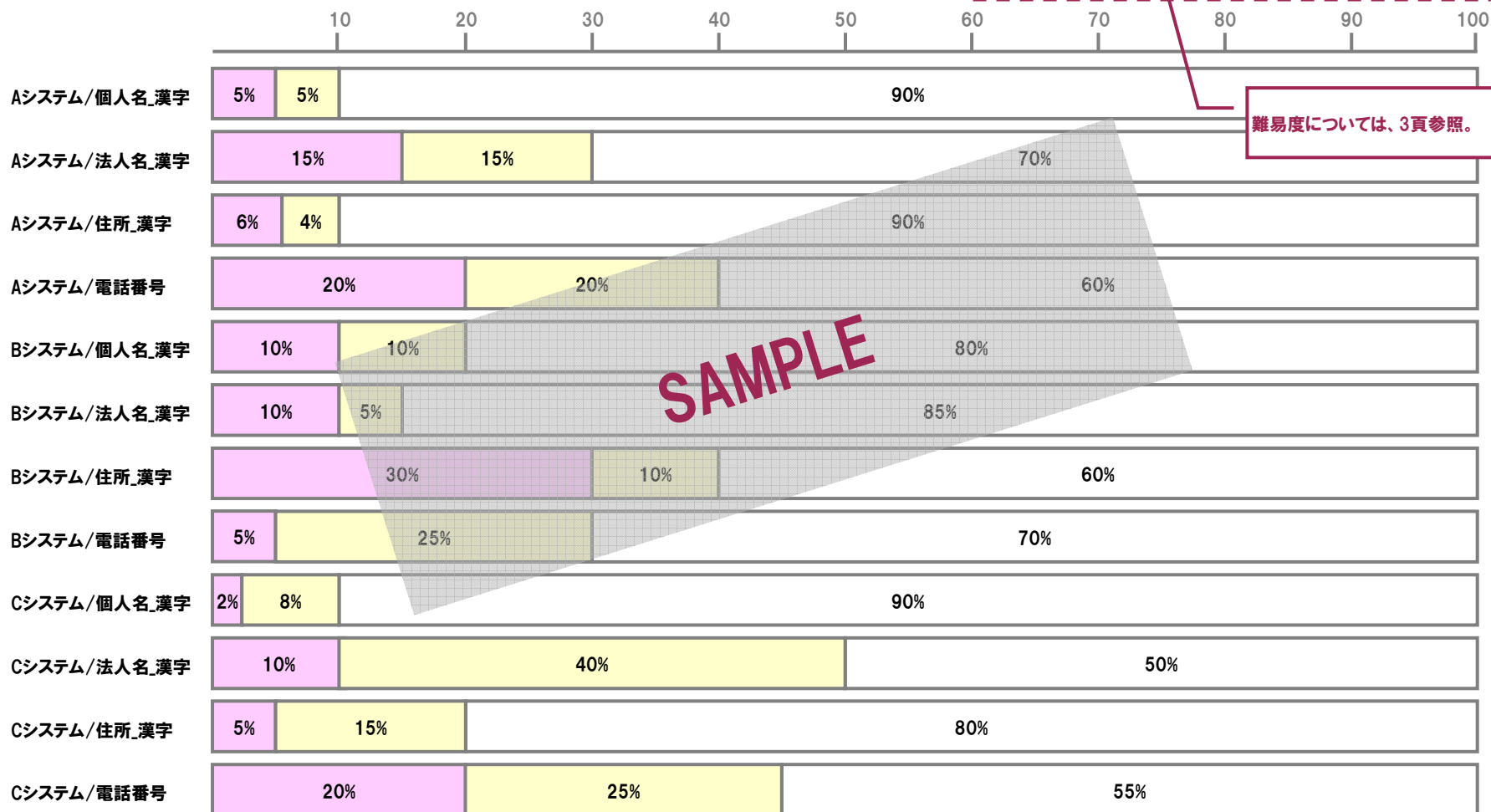
診断概要

・Cシステムは問題データの割合がカラム毎にバラツキがあるが、全般的にカテゴリ3.の割合が多く、ツールによる補正が効果的であると考えられる。

・「個人名_漢字」カラムはシステムに因らず全般的に問題データの割合が少ない。

...

■ : 問題データ(難易度高) ■ : 問題データ(難易度低)
□ : 有効データ



難易度については、3頁参照。

SAMPLE

Quickプレ診断書_詳細

システム毎/カラム毎に、主な問題データの件数/例、それら問題データが生成される想定原因、それら問題データが存在することによる業務への影響、それら問題データを解決するための推奨解決策などをまとめた診断結果。1頁のグラフ単位に作成される。

□対象データ

システム	Aシステム	カラム論理名	個人名_漢字
データベース	Oracle 10g		
型	VARCHAR2	桁	20桁
受領ファイル形式	TXT.ファイル	ベースとなる文字コード	Shift JIS
受領レコード件数	500,000件	調査レコード件数	50,000件(Sequence No.末尾が0のレコードを抽出)

□問題データ診断結果

カテゴリ	件数	割合	主な問題データ	想定原因/業務への影響/推奨解決策等
難易度高	2,500件	5%	NULL値:500件 空白:200件	<ul style="list-style-type: none"> これら難易度高の問題データは、証拠に遡って検証する以外に解決方法はありません。 これら難易度高の問題データは、割合は5%と少ないですが件数は2,500件であり、見過ごすことはできないと考えます。難易度低の問題データと比較して解決に工数がかかりますが、「ありたい姿」の実現に向け、証拠に遡って解決されることを推奨します。 NULL値や「●」、「■」等の記号は、入力ミス以外の原因として、過去のデータ移行に於ける変換ミスや、異なる環境間をデータ転送された際に環境依存文字が対応せずに変換された可能性があります。データが生成されてから活用に至るまでのデータの流れを整理し、原因の解明とこれらのデータが生成されないための歯止めのプランニングと共に、再入力をご推奨します。 個人名_漢字カラムに法人名が混入されていることは、入力ルールが整備されていない、若しくは遵守されていないことが最大の原因であると考えられます。原因の解明とこれらのデータが生成されないための歯止めのプランニングと共に、再入力を推奨します。 個人名_漢字カラムは顧客を特定する上で最も重要な項目であり、「あいうえお」のように顧客の特定が完全に不可能なデータは本来存在してはなりません。これらのデータは過去のデータ移行やデータ転送、入力ミスに起因するものではなく、人間により意図的に入力/生成された可能性が高いデータです。本診断の中では30件と微小ですが、絶対に存在してはならないデータですので、原因の解明とこれらのデータが生成されないための歯止めのプランニングと共に、再入力を推奨します。
			ルール不適合:100件 (株)リアライズ ・ススキイチロウ 不明:30件 ・あいうえお ・●藤太郎 ・■橋花子	
難易度低	2,500件	5%	異体字:100件	<ul style="list-style-type: none"> これら難易度低の問題データは、データとしては間違っているものの近似の値として判断することは可能であり、また一般的なツールに於いて辞書定義/変換定義を保持しているものです。 難易度高の問題データと同様に、このままではダイレクトメール不達、分析精度の低下等の事象を引き起こしますが、難易度高の問題データ程はクリティカルな問題ではないと考えます。 データクレンジング、名寄せを実行する適切なツール選定を支援しますので、ツールで実施可能な範囲で解決されることを推奨します。
			類似:20件	

難易度については、3頁参照。

□問題データのカテゴリ

難易度 高	○原本まで遡って確認/再入力することで解決できる問題データ <ul style="list-style-type: none">・氏名が「あいうえお」等、原本を確認することによって真の氏名を判別できるもの。・生年月日が「明治01年」等、原本を確認することによって真の生年月日を判別できるもの。・氏名なのに「19710508」等、原本を確認することによって真の氏名を判断できるもの。
	○人間が判断して定義した上で、ツールを用いて解決できる問題データ <ul style="list-style-type: none">・「唱和」等、入力ミスが起こりやすいが、人的にパターンを検出しなければ、ツール処理が困難なもの。・「有)」「有限)」「(ユウ)」「(株)」「(株)」など、人的にパターンを検出することで、ツール処理がより効果的となるもの。・「アベ/ABE」等、ツールによる検出と人的な判断を合せて行う必要があるもの。
難易度 低	○ツールで自動的に解決できる問題データ <ul style="list-style-type: none">・「斉藤/齋藤」や「シヅコ/シズコ」等、ツールに於いて辞書定義されているもの。・「和暦⇒西暦」等、ツールに於いて変換定義されているもの。